

LEVEL III

A064272
VOL III

ARI TECHNICAL REPORT
TR-78-A27 (diff m)

**Improved Army Training and Evaluation Program
(ARTEP) Methods for Unit Evaluation
VOLUME II: ANALYSIS**

by

M. Dean Havron, Doris D. Albert &
Timothy J. McCullough

HUMAN SCIENCES RESEARCH, INC.
7710 Old Springhouse Road
McLean, Virginia 22101

and

Robert G. Wanschura
U.S. Army Research Institute for the
Behavioral and Social Sciences

NOVEMBER 1978

Contract DAHC 19-77-C-0001
Project: PR No. 76-68-PERI-OU
Engagement Simulation Technical Area

Prepared for



U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333



79 04 03 036

AD A0 66783

DDC FILE COPY

**Best
Available
Copy**

**U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES**

**A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel**

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, US Army
Commander

Research accomplished
under contract to the Department of the Army

Human Sciences Research, Inc.

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ART. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1 TR-78-A27	2. GOVT ACCESSION NO. 18 ARI	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPROVED ARTEP METHODS FOR UNIT EVALUATION: VOLUME II: ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report 8 Nov 76 - 31 Jan 78
6. AUTHOR(s) 10 M. Dean Hayron, Doris D. Albert, Timothy J. McCullough and Robert G. Wanschura (ARI)		7. PERFORMING ORG. REPORT NUMBER HSR-RR-77-21-Ge
8. PERFORMING ORGANIZATION NAME AND ADDRESS Human Sciences Research, Inc. 7710 Old Springhouse Road McLean, Virginia 22101		9. CONTRACT OR GRANT NUMBER(s) 15 DAHC 19-77-C-0001
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Training and Doctrine Command Fort Monroe, Va. 23651		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 2Q263743A773
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) U.S. Army Research Institute for the Behavioral & Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333		12. REPORT DATE 11 November 1978
		13. NUMBER OF PAGES 57
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited 12 62 p. 14 HSR-RR-77/21-GE-VOL-2		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) 6 Improved Army Training and Evaluation Program (ARTEP) Methods for Unit Evaluation. Volume II. Analysis.		
18. SUPPLEMENTARY NOTES Research monitored technically by Robert G. Wanschura, Engagement Simulation Technical Area, Army Research Institute.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Evaluation Performance Army Training and Evaluation Program (ARTEP) Tank/Mechanized Infantry Task Force		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This study analyzes existing methods of implementing the Army Training and Evaluation Program (ARTEP) for a Tank/Mechanized Infantry Task Force. Principles of the ARTEP system need to be more fully developed before field units can use them to develop training and evaluation programs that effec- tively meet the goals of the ARTEP concept. Battalion field evaluation exercises should be planned with the ultimate goals of training diagnosis in mind. Therefore, significant emphasis must be placed on adequate		

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Block 20 (Continued)

training for evaluator/controllers. A three-volume report provides guidelines for meeting these requirements.

The report is intended primarily for readers interested in collective training, particularly those working with the development of ARTEPs.

Volume I is of general interest to ARTEP developers, training managers, policy makers, and users. Volume II provides data analysis and recommendations for refining current ARTEP implementation; it will interest developers, training managers, and policy makers. Volume III is a prototype guide for battalion-level use of the ARTEP. It is based on the recommendations from Volume II and is of interest to senior commanders, their staff, and officers who train others to perform evaluations and exercise control functions in the field.

ACCESSION for	
NTIS	While Section <input checked="" type="checkbox"/>
DOC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	<input type="checkbox"/>
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	Avail.
A	

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

CONTENTS OF THIS VOLUME

PART TWO: ANALYSIS	1
Chapter 1: ARTEP System Concepts and Major Applications	2
1-1. Introduction	2
1-2. ARTEP System Concept Issues	2
1-3. Issues, Application of ARTEP Concepts to the Evaluation Process	5
1-4. Preparing and Implementing the Evaluation Plan	11
Chapter 2: Evaluation Technical Methodology	22
2-1. Introduction	22
2-2. Guiding Concepts	23
2-3. Use of Current T&E Outlines as Evaluation Instruments	34
Chapter 3: Use of Evaluation Results	45
3-1. Introduction	45
3-2. Formulating and Communicating Evaluation Results	47
3-3. On-Line and Written Feedback	51
3-4. Use of Results	53
Chapter 4: Summary of Major Results and Recommendations	55
4-1. Introduction	55
4-2. ARTEP Systems Concepts	55
4-3. Evaluation Technical Methodology	55
4-4. Use of Evaluation Results	56
4-5. Recommendations for Future Research	56
4-6. The Field Guide	57

79 04 03 036

PART TWO: ANALYSIS

CHAPTER 1

ARTEP SYSTEM CONCEPTS AND MAJOR APPLICATIONS

1-1. INTRODUCTION

This volume presents the analysis for the first project year. The chapters parallel the major issues introduced in Volume I. These issues are reproduced on page 15. This chapter deals with ARTEP management and the preparation of Tank/Mechanized Infantry Task Force evaluations. Chapter 2 focuses on evaluation methodology. Chapter 3 considers issues in the use of data obtained from battalion evaluations. Chapter 4 summarizes this volume. Each subject chapter:

- Provides a general background discussion of the issues, and approach to their resolution.
- Distinguishes between issues with preliminary "solutions" for the Draft Field Guide, and those which require further research.
- Summarizes the analyses leading to the guidance included in the Draft Field Guide.¹
- Recommends follow-on studies for issues which could not be solved in the first year.

1-2. ARTEP SYSTEM CONCEPT ISSUES

a. **Background.** This chapter deals with principles of ARTEP management and with trade-off decisions that must be made in applying them to evaluation exercises. Many problems with the application of ARTEP stem from failures to make these decisions. These failures, in turn, derive from the current state of development of the ARTEP system and its core principles. There is no single source document which pulls them all together and focuses them on ARTEP evaluation procedures.

The best statement for putting these notions into field practice is found in the two versions of ARTEP 71-2, Chapter 4. The draft version includes a detailed sequence of ten steps for training program development, drawn from TC 21-5-1 and TC 21-5-2. The later version follows TC 21-5-7 by combining certain of these detailed steps. Their place is taken by a four category flow-chart (Analyze, Provide, Conduct, Evaluate), and a substantial amount of helpful but still less than complete coverage of what brigade/battalion commanders and S-3's seem to need. TC 21-5-7 is cited as the key complementary reference, which supplies further explanation and illustration of how the basic principles are to be understood and used.

¹ The Field Guide appears in this report as Volume III. The Guide is organized into "Introductory," "Senior Command and Staff," and "Evaluator/Controller Group" modules.

The single master concept of ARTEP is of course "performance oriented training." Inspection of definitions and use of this concept in ARTEP 71-2, FM 21-6, and TC 21-5-7, reveal substantial variation in its meaning, and dependency for sense on an array of more or less closely associated ideas, including:

- Mission/task oriented training/evaluation.
- Concurrent, multi-echelon training/evaluation.
- Training/evaluation to correct deficiencies.
- Decentralized training/evaluation, etc.

In addition, some twelve "underlying principles for sound training management," which complement and overlap the above, appear in the same documents. Among these are "austerity," "readiness," "realism," "accountability," and "command emphasis" (on training missions), among others. Roughly parallel usages occur in the corresponding FORSCOM regulations and guidelines.

The field observations indicate significant areas of difficulty in determining what the basic ARTEP system concepts are: how they relate to each other and how they differ from earlier training/evaluation tools. Thus, it is difficult to determine how they must be employed by units to get the full benefits from ARTEP.

Examples from a more extensive list² which illustrate the problem are:

- How to interpret and use performance standards provided in the T&E outlines.
- How to balance and schedule a sound mix of training and evaluation activities of all types in relation to other demands on Division/Brigade/Battalion resources.
- How to determine and allocate responsibilities for implementation of ARTEP "decentralized" training/evaluation principle.
- How cycling and "peaking" tendencies can be successfully replaced by sustained "concurrent training," and how such concurrent training can sustain high levels of proficiency and "readiness."

²For a fuller description, see *Final Report (DRAFT). Improving Army Training and Evaluation Programs (ARTEP); Methods for Unit Evaluation-- Volume I. Executive Summary: Study Design and Field Research.* 31 January 1978. Chapter 3.

- How requirements for "realistic" training/evaluation can be met within the constraints of "austere" use of available assets.
- How external evaluations administered by higher headquarters can still be made to retain "diagnostic" rather than "test" characteristics.

b. **Separation of Near-Term and Longer Term Issues.** Certain aspects of these issues do not lend themselves to ready solutions. Such solutions would involve basic program development efforts that were outside the scope of Phase I research. Therefore, the ARTEP core principles were accepted in their current form and near-term quick fix remedies were developed. These solutions are recommended in Volume III, page B-4 to B-8. These stop-gap measures involve collectively applying the ARTEP core ideas and training management principles by making practical trade-offs among them. These measures are considered next.

c. **Field Guide Remedies.** Examination of sources yielded a list of some fifteen major concepts as candidates for treatment as elements in the "system of core ARTEP principles." Given the need for brevity in the Field Guide, this set was reduced to seven essential concepts:

- performance orientation
- decentralization
- concurrent, multi-echelon approach
- inseparability of training and evaluation
- realism
- austerity
- accountability

The Field Guide was developed so as to be compatible with guidance pertaining to the above concepts. It consists of two modules: one for Senior Commanders and Staff, and a second for Evaluator/Controllers. Each of the above concepts is discussed in the Field Guide *Senior Command/Staff Module* as it bears on evaluation field practice. Their interdependence is stressed. Finally, several examples of reasoning and procedures involved in finding practical trade-offs among the principles are provided.

d. **Recommendations for Follow-On Research.** Longer term solutions to ARTEP problems require further refinement and development of ARTEP as a training management system, and possibly the integration of new technologies into its evaluation component (e.g., engagement simulation, battle simulation, etc.). This "master plan for guiding ARTEP system development and implementation" might include three tasks:

- (1) Clarification and projection of core ARTEP program objectives.
- (2) Identification and prioritization of problem areas.

- (3) Definition and sketch design of projects which require managerial and/or research attention.

Based on our observations and analysis, we suggest three candidate projects:

- Analysis of the performance orientation model as applied to collective training of large units. This model applies quite well to "hands-on" tasks. As a minimum, it needs to be elaborated for comprehensive application to data processing tasks of tactical decision-makers. Precise, objective criteria are often difficult to formulate and to apply with validity at these levels.
- Analysis of the test/diagnosis issue area to identify the major available options to Senior Commanders and field planners. These personnel need concepts and tools for *concurrently* monitoring readiness and accountability for diagnosis of remedial training needs.
- Analysis of current problems involved in defining and applying sound T&E standards for large units. This effort should help to establish better understanding of when and how performance can be measured by objective standards, and when and under what conditions application of criteria must depend, for the most part, on professional judgments.

1-3. ISSUES, APPLICATION OF ARTEP CONCEPTS TO THE EVALUATION PROCESS

A number of concrete questions need to be answered in order to apply the ARTEP core concepts to field evaluations. Specific issues include:

- Deciding to conduct an external or internal evaluation.
- Designating key responsibilities.
- Deciding how amounts of assets are committed.
- Guidance needed to "keep the focus on diagnosis."

Each is discussed separately below.

a. External or Internal Evaluation?

(1) Background discussion. TRADOC/FORSCOM guidelines provide two alternatives for evaluating units. The internal approach uses inexpensive, informal means to yield rough

diagnostic data. The external evaluation employs costlier and more rigorous means to produce systematic data for both unit personnel and senior commanders. It requires a more carefully timed, less frequently scheduled effort, which must be integrated into brigade/division timing schedules. Current practice of administering an external evaluation every 12-18 months perhaps serves as a realistic compromise between what is desired and what is possible. But when in that period should the external evaluation be scheduled?

(2) Separation of the Field Guide and long-range solutions. Current guidance to schedule external evaluations by "locating them in the unit's total T&E program may be asking too much of local commanders. In order to do that systematically, each commander would require a management tool which allowed him to array, at least roughly, the full range of training and evaluation procedures and techniques available to him. He would also require some basis for weighing these options for their comparative costs and benefits in relation to recent T&E activity, current proficiency levels, and proficiency targets. These management aids do not currently exist in any but the most rudimentary form. Moreover, development of such training management tools, allowing commanders to approach the optimal mix of options for their commands, lay substantially beyond the scope of Phase I work for the present study.

Phase I analysis attempted to clarify the conditions under which a specific battalion could best use the external evaluation option. This in turn led to consideration of measures which would allow senior commanders to "time" external evaluations so that they would begin to mesh better with the readiness of individual battalions in their command at specific points in time. These points are briefly considered next.

(3) Field Guide remedies. The Phase I guidance developed appropriate criteria for the adjustment of training schedules. The scheduling recommendations took note of the very limited flexibility typically available in unit schedules. It concentrated on *forecasting* evaluation preparedness and on building such forecasts into long-term master T&E schedules. This material is incorporated in the Senior Command/Staff Module of the Field Guide.

b. Designation of Key Responsibilities

(1) Background discussion. One of the core principles of the ARTEP system is that, all else being equal, the more decentralized the training/evaluation process:

- the more experience, sense of involvement and responsibility, and professional development lower echelon leaders will enjoy.
- the better training/evaluation can be tailored to fit the particular needs of the units involved.
- the more accustomed leaders will be to control of dispersed elements on modern battlefields.

On the other hand, an ARTEP principle just as strongly emphasized as decentralization is that of accountability. Reconciliation of the two becomes a practical need when decisions must be made about where to locate major task assignments for the evaluation effort. The basic question is how to find a reasonable balance point between the two.

(2) Separation of the Field Guide from long-range solutions. Finding a good balance between such principles as decentralization and accountability may become unnecessary if studies point toward the use of separate instruments for them. For example, the best approach may be to employ ARTEP evaluations exclusively for diagnostic purposes, and to introduce other procedures to assess readiness and leader accountability. Short-run remedies required inspection of the major "pro's" and "con's" that characterized brigade sponsorship and control of the evaluation effort versus that by division.

(3) Field Guide remedies. Existing guidance in this area, to be found in Chapter 4 and 5 of ARTEP 71-2, offers only generalities about an appropriate division of labor. The key issue of whether to locate the primary center of activity at the division or the brigade headquarters level is not resolved. Although roles are identified in Chapter 5 for a "chief evaluator," an "officer in charge," and a "senior evaluator," it is not possible to decide to what degree these titles are intended as synonymous. More importantly, it is unclear whether one or more of these is intended to coincide with the role of the "senior commander (brigade and higher)" whose broad training and evaluation responsibilities are outlined in Chapter 4.

The substance of the existing guidance seems, therefore, to be that either the brigade or division commander (or his designee) may operate as chief evaluator or responsible senior commander. As a corollary, the location of the recommended "ARTEP evaluation group which prepares and conducts evaluations" might be at either corresponding staff level. But which pattern of predominance, division or brigade, is preferable where a choice is possible?

In favor of the division-centralized approach are several factors. These include greater control over more resources, the possibility of developing greater technical expertise by a semi-permanent and specialized staff group, and certain potential economies in the use of staff time to do the basic job once for all battalions rather than having each brigade duplicate the efforts of every other brigade in this area. In addition, this approach enjoys the advantage of potentially very powerful "command emphasis," which differs in kind when emanating from the division commander rather than a brigade headquarters.

The advantages of the brigade-centered approach seem more impressive. One key relative advantage, of course, is that brigade is an echelon closer to the battalion(s) to be evaluated. This meets the minimum requirement for externality, while at the same time promoting the previously cited virtues of a relatively decentralized approach. In addition, this formula avoids many of the liabilities that go hand in hand with the more attractive features of division-centered evaluations. These include:

- A lessened tendency to overspend (because fewer assets are present and "borrowing" requires formal approval and justification).

- A reduced potential for misperception of the evaluation as an "annual test" imposed by a remote command echelon.
- Less likelihood that the scenario/evaluation plan will become rigidly stylized, "canned," devoid of surprises or unanticipated tactical events, etc.
- More likelihood that the evaluation will be closely tailored to the specific requirements of the individual brigade and its battalions.
- Less likelihood of delays and distortions in the feedback process, with the result of more timely and appropriate changes in unit training activity.

On the basis of these considerations and available evidence, a conclusive preference for either basic option is unwarranted. The most that can be said is that the edge seems to lie with the brigade-centered alternative, although this edge derives largely from the attachment of subjective weights to the several criteria involved. Accordingly, guidance addressed to the senior commander (division or brigade), is framed to include the following elements:

- Explicit recognition that multiple advantages and disadvantages are present in both the division-centered and brigade-centered approaches.
- A brief sketch of what these are, as summarized above.
- A qualified assessment that the brigade-centered approach may be a better overall compromise in relation to fundamental ARTEP principles, except in unusual circumstances (e.g., a decision to use an "integrated ARTEP" or two-battalion evaluation format, as considered in the next section).
- A strong recommendation to the effect that, whichever alternative approach is adopted, this same focal point of evaluation activity should be retained and used consistently throughout each of the preparatory, conduct and results utilization phases of the evaluation program as a whole.

In addition, guidance goes on to develop moderately detailed recommendations concerning the composition and activities of the "ARTEP Evaluation Planning/Writing Team."

(4) Recommendations for follow-on research. Final resolution of the issue as to whether the brigade or division-centered approach is better will require a rigorous comparative analysis of the relative costs and benefits associated with these options.

c. Deciding How Many Assets to Commit.

(1) Costs must be calculated in terms of some notion of the resource total, and in relation to "opportunity costs."³ By opportunity costs, we mean those other activities which must be foregone or pursued less intensively because of commitment of resources to battalion evaluations.

The danger is either to over- or under-emphasize support of the evaluation effort. Relatively sparse support in some units risked serious compromises in tactical realism and evaluation validity. On the other hand, some sponsors, we felt, over-committed resources for the evaluation field exercise, with resulting "VIP" or "dog and pony show" overtones, a high-pressure rather than diagnostic atmosphere, and presumably heavy "opportunity costs" elsewhere in the training/evaluation equation.

(2) Separation of near-term and follow-on remedial work. Definitive guidance on how many assets to commit will require further experimentation with respect to the full array of training/evaluation options (e.g., CATTs/CAMMs Simulations, CPX, TEW, FTX, differing procedures at differing echelons, etc.). For the near-term, however, Phase I guidance stresses the greater likelihood of underspending, and how reducing the frequency of external evaluations offsets higher costs.

(3) Field Guide remedies. Our field observations and comparison of cases suggest that underspending is more likely than overspending. Therefore, the basic recommendation on this issue is to be prepared to spend more on individual exercises and to use what is spent more efficiently. To offset the resulting higher asset costs, it is suggested that senior commanders consider stretching out the interval between external evaluation exercises toward the upper limit figure of 18 months. In conjunction with less frequent external evaluations, more use of internal evaluations is also encouraged.

(4) Recommendations for follow-on research. As a minimal target, some type of accounting scheme should be developed and fielded which makes possible the uniform calculation of costs involved in external evaluations (and for that matter, for all aspects of units' T&E programs). If this tool and resulting data were available, definitive studies of the relative payoffs of differing support levels could be executed, and more explicit guidelines derived.

d. Selecting Measures to "Keep the Focus on Diagnosis"

(1) Background discussion. Perhaps the single most widely acknowledged problem in the entire ARTEP evaluation process is that of successfully upholding the definition of the evaluation as a diagnostic tool, rather than as some version of a commander's report card or unit proficiency/readiness test. The basic distinction is quite clearcut. The underlying ARTEP doctrinal principles are equally explicit. And yet the problem persists.

³At present there is no tidy basis for "costing out" external evaluations for comparisons of any sort between units or parent organizations.

Part of the explanation no doubt lies in the area of "institutional inertia," or ingrained habits which confront any major innovation, and as such presumably respond to vigorous educational efforts. Battalion field evaluations are very expensive. For this reason it is difficult to completely eliminate the element of accountability. Given this, officers evaluated are prone to try to stick precisely to doctrine rather than to be innovative. Further, their orientation tends toward rationalization of their acting, rather than learning and diagnosis of deficiencies. The fact is that at present performance in field evaluations provides the best available means of holding commanders and training managers accountable. It follows that this view will not be changed by even the most emphatic declarations to the contrary, if unsupported by concrete changes in the way evaluations are defined and used.

(2) Separation of near-term and follow-on remedial work. Resolving the contradictory requirements of readiness assessment, accountability and training diagnosis go beyond the scope of this project. Phase I guidance suggests to senior commanders actions to be taken to keep emphasis on training and diagnosis.

(3) Field Guide remedies. The immediate need was to provide effective short-run means for containing "test" tendencies sufficiently to prevent their undermining of diagnosis and correction of training deficiencies by the people in the best position to do so.

Several factors which promote the "test" perception are subject to control. These include:

- Division level command/staff predominance in all phases of the effort, overly "externalizes" the process and pressure on battalions to "look good."
- Adoption of the "opposing forces" or "integrated ARTEP" format in which two battalions are evaluated *against* each other.
- Lack of provision in the training schedule for the battalion and its commander to demonstrate deficiencies diagnosed have been remedied. This one-shot aspect inevitably heightens the incentive to "look good" the first time, rather than to be completely open to accurate appraisal of deficiencies.

Taking these factors into account, short-run remedies were given the form of a seven point procedural package the senior commander could employ. The most critical element is recommendation of a paragraph to be included in the basic LOI, which underscores the diagnostic rather than "report card" nature of the evaluations, and explicitly disclaims any intent to attach sanctions to the results.

In addition, attention is called to each of the exacerbating factors cited above, with the recommendation that evaluation planning, conduct and utilization be adjusted to assure a "sanction-free" learning environment.

Equivalent treatment of the same basic concepts for promoting the diagnostic nature of the ARTEP exercise is also prominently displayed in the Evaluator/Controller Group Module.

(4) Recommendations for follow-on research. The impact of the package recommended above may be assessed when the first generation Field Guide is evaluated in Phase II.

1-4. PREPARING AND IMPLEMENTING THE EVALUATION PLAN

Issues addressed in this section apply to preparation of the master Evaluation Plan and its field implementation. Issues dealing explicitly with evaluation methodology, have been set aside for separate treatment in the following chapter. This section focuses on several problems every field unit must resolve. These include:

- Choosing a basic exercise format.
- Constructing an adequate scenario.
- Providing effective exercise control and tactical simulation procedures.

a. **Choosing an Evaluation Exercise Format.** Several choices to be made by senior commanders and planners regarding format and procedures are fundamental. Every subsequent planning, conduct and feedback step is at least partly foreshadowed by the particular option exercised. Three have been selected for treatment here.

(1) How many units of what types should participate in what roles?

(a) Background discussion. Which units of what types should the senior commander decide to include in the evaluation exercise, and with what assigned roles? This broad question establishes the basic terms for construction of the details of an overall Evaluation Plan. It may be broken into two parts. The more critical question is whether to conduct the evaluation for a single battalion, or whether to attempt to "integrate" the evaluation of two "opposing" battalions into a single exercise. The other aspect concerns which elements to use in assembling the battalion task force(s), and what units to incorporate in supporting roles.⁴

On the one or two battalion issue, current FORSCOM regulations explicitly state, "Simultaneous evaluation by more than one maneuver battalion is not recommended."⁵ Despite this, our observations confirm that the simultaneous format continues to be used extensively in FORSCOM units. But what are the relative merits of the single or integrated approaches?

⁴To save space, these latter relatively simple issues will be bypassed with the notation that the universal practice in forming Task Forces is by cross attachment of single company sized units. Patterns of commitment of supporting units generally follow the recommendation of providing "a battalion's fair slice of division assets" of artillery, engineers, signal, helicopters, tactical air, etc.

⁵FORSCOM REGULATION 350-1, Chapter 3, Paragraph 3-4.e.

Proponents of the integrated approach typically assert the following advantages:

- It is more cost effective, because it requires only one preparatory effort, one evaluation/control group, one period for use of maneuver areas, and no dedicated aggressor force, among other possible economies.
- It is more realistic, because it puts more realistic force ratios against each battalion, and generally injects an element of unpredictability into the exercise due to the degree of tactical freedom exercised by each.
- It produces higher levels of involvement, enthusiasm, performance, and training value as a result of the more competitive atmosphere.

Proponents of the single battalion evaluation argue these key points:

- It is more cost effective, because it reduces the temptation to make massive use of assets, particularly those drawn from outside the battalion.
- It requires proportionally *fewer* control/evaluation personnel than the other approach, because problem control and evaluation procedures are significantly simplified.
- It provides more realism, because the dedicated aggressor force may be played with authentic threat doctrine and configuration, unlike in the other format.
- It is a better evaluative vehicle inherently, because the degree of "stimulus control" required to conduct refined measurement of behaviors is very much higher. The OPFOR is a manipulable instrument.
- There is less likelihood that the division command/staff level will be drawn into a predominant role, thus compromising the previously cited advantages of decentralization.
- The relative simplicity of managing a field exercise involving a single battalion makes fewer demands on the ARTEP experience of planners/evaluators/controllers, and minimizes distractions from core evaluation objectives.
- The relative simplicity and controllability of the tactical situation makes the scheduling and conduct of all types of feedback and on-line training easier.

(b) Field Guide remedies. Our analysis of the relative merits of both alternatives led to a recommendation in the Senior Command/Staff Module to use the single battalion approach, for several reasons. First, there is little gain in realism involved in fighting another U.S. battalion, and many obvious artificialities. Also, the problems of successfully controlling, evaluating and providing adequate tactical simulation for two large units simultaneously are almost unsolvable within reasonable resource limits. And finally, the planning, coordination and support requirements for the two-battalion approach tend to award the sponsorship function to division by default. This impinges on the advantages of decentralization and diagnosis cited in the previous example. Accordingly, the better format appears clearly to be the conventional one-battalion exercise approach.

(2) Where to place evaluation priorities and emphasis?

(a) Background discussion. The major concern here is how to organize the evaluation process and distribute evaluation resources in relation to priorities among the various units and echelons composing the battalion task force. Several alternative patterns of prioritizing are possible. However, little guidance is currently supplied and substantial differences in evaluation results depend on the choices made.

In the field at present, if the way evaluators are distributed within the battalion is taken as a rough index of emphasis, it is the battalion rather than company or platoon levels which receive by far the greater share.⁶ This amounts to saying that data collection on the battalion level is more valuable, or is required in greater amounts, or both. It is easy enough to understand how in a battalion evaluation a disproportion of emphasis gets placed on the *highest organizational plane*. However, there are no compelling reasons for doing this, and several important ones for spending more evaluator assets at the lower unit echelons.

One such reason is that the bulk of the behavior which requires observation and evaluation takes place in the maneuver companies. This is true in the sense that the sheer bulk of T&EO items aggregate for the three companies is significantly greater than for the battalion. It is also true that a large proportion of the battalion items represent no observable behavior at that echelon, but rather consist of summations of observations made at the company and platoon levels.

Finally, it is worth recognizing that conduct of valid evaluations increases in difficulty with the complexity of the units involved. At present, the "state-of-the-art" with regard to collecting highly meaningful and usable data on larger-sized units is still fairly primitive.⁷ It is arguable that these limitations can be bypassed by placing less emphasis on the battalion as the primary focus for evaluation, even in a battalion formal evaluation. In this way, more evaluators can be concentrated at echelons where the data validity problems are less intimidating.

⁶Among our cases, the typical battalion evaluator team is composed of roughly 25 personnel, of whom only 2 or 3 are with each maneuver company, aggregating to less than half the total.

⁷The question of whether battalion scale evaluations can be justified under current circumstances, given their high costs and relatively limited results, is raised explicitly in the next chapter.

(b) Field Guide remedies. In line with these points, the E/C Group Module incorporates explicit suggestions for achieving a better balance and evaluation coverage among the several unit levels. It is recommended that no less than half the total battalion evaluator team be employed at the company and platoon levels. Details of evaluator/controller team organization are weighed below, but on the assumption of a 25-28 man team, the formula recommended here would permit the use of about 15 evaluators with the maneuver companies. See Figure 1-2, page 20.

(3) If, when and how to conduct "sub-unit" evaluations?

(a) Background discussion. The essence of this issue is how to conduct "sub-unit" evaluations" which yield valid performance data at minimum cost to the larger evaluation process. These sub-unit evaluations can incur such costs by drawing off evaluator personnel, diverting units from the battalion problem, fatiguing units before the battalion problem, and so on.

If company level evaluations are excluded from this category, the ARTEP 71-2 sub-unit missions included in Chapter 8 fall into two basic types. One of these involves the sub-unit in behaviors which are *highly sensitive to the context of the larger unit and tactical situation*. In this case, removal of the unit (e.g., Tank/Mechanized Infantry Platoon: Defense), from the battalion scenario significantly dislocates the platoon from a realistic setting, while at the same time significantly impacting the larger problem from which the platoon is removed. The other case—in which *the sub-unit mission is not highly sensitive to a larger tactical context* (e.g., Tank Platoon Battle Run), or where special range or safety requirements make integration into the battalion exercise unwieldy—presents a situation where the preferred choice is to conduct the mission evaluations "off-line."

(b) Field Guide remedies. For those sub-unit evaluations to be conducted off-line, the question of when off-line evaluations are to be conducted in relation to the battalion exercise must be answered. Except where range access constraints mandate it, conducting these sub-unit evaluation simultaneous to the battalion evaluation should be avoided due to its disruptive effects on parent units and the evaluator team. Conducting sub-unit evaluations at the end of the battalion exercise also has serious problems that result from personnel fatigue and the delays entailed in scoring the sub-unit evaluations and in integrating these scores into the larger unit's feedback procedures. The best all-around solution, therefore, is to conduct sub-unit evaluations *before* the battalion exercise. In this case, the one key liability—that of wearing out the sub-unit personnel just before the 3 to 4-day battalion problem—can be handled by scheduling the sub-unit evaluations to permit a 2 to 3-day break before the battalion problem begins.

To sum up, the recommendations incorporated in the Draft Guide on this issue were:

- describing the distinction between context-dependent and other sub-unit missions.
- recommendation that only context-dependent missions be conducted on line.

- recommendation that all other sub-unit evaluations be conducted prior to the battalion evaluation (except where range or similar access problems mandate otherwise), allowing several days after these sub-unit evaluations for troop rest and maintenance before the larger exercise.

b. Scenario Construction. The heart of the overall Evaluation Plan is the basic exercise scenario, which coordinates all administrative, tactical and evaluation activities in a single procedural sequence. This section considers the major issues involved in sound scenario construction. We will consider first how many and which missions to represent, and then the issue of sequencing these appropriately.

(1) How many missions?

(a) Background discussion. We note initially that ARTEP 71-2 mandates that at least six of the nine basic battalion missions, and seven of the eleven battalion "supplemental" missions be evaluated. Our field observations indicate that seven or eight basic missions and eight or nine supplemental missions is current practice.

Our observations also indicate that the temptation to "get our money's worth" by squeezing too many missions into the exercise creates serious problems. Among these problems are:

- A rapid succession of basic missions is not likely on the modern battlefield, and, therefore, undermines tactical realism.
- Too many basic missions unrealistically compresses the time frame in which all facets of unit behavior occur. This time compression is especially damaging with respect to permitting adequate time for "troop leading procedures" at company and lower echelons.
- The larger the number of basic missions, the more difficult it becomes to provide in an orderly fashion for timely critique and other types of feedback.
- The larger the number of basic missions, the more complicated administrative, tactical and evaluation control procedures become, leading to greater likelihood of breakdowns.
- The larger the number of basic missions, the less feasible becomes the already highly complicated process of conducting an "integrated" or two-battalion evaluation.

(b) Field Guide remedies. For all of the above reasons, it appears highly appropriate to encourage senior commanders and planners to stick to the minimum or near the minimum in the number of basic and supplemental missions they attempt to evaluate in one 3-4 day exercise. If the unit is inexperienced in the conduct of battalion external evaluations, or if the "integrated"

format is selected, such restrictions are almost a precondition for a successful effort. Accordingly, we strongly recommend in the Draft Guide that six, or at most seven, primary missions and seven, or at most eight, supplemental missions be attempted.

(2) Which missions?

(a) Background discussion. The basic battalion missions in ARTEP 71-2 are:

- Movement to Contact
- Hasty Attack
- Deliberate Attack
- Exploitation
- Night Attack
- Defense
- Delay (High Risk)
- Disengage (Under Pressure)
- Defense of a Built-Up Area

Of these, the Defense of a Built-Up Area is unlikely to be practicable because almost no training areas provide adequate facilities. Similarly, the deliberate attack may be regarded as impracticable because of the prohibitive amount of time required in relation to the 3-4 day total. This leaves only seven basic missions, for which we find no compelling reason to recommend some rather than others. We recognize, however, that another view is possible, asserting that some subset of these (say Movement to Contact, Hasty Attack, Defense, Delay) represents an irreducible core for a combat ready unit, and therefore, *must be used* in the external evaluation. Such assertions embody doctrinal assumptions that would require a thorough airing before becoming formal guidance. For the same reasons, we are reluctant to attempt to select from among the eleven supplemental missions provided in Chapter 9 of ARTEP 71-2.

(b) Field Guide remedies. Phase I field guidance leaves planners free to use their professional judgment and other appropriate considerations in choosing among the basic and supplemental battalion missions.

With respect to sub-unit mission evaluations, a similar line of reasoning leading to recommendations to stay close to minimum requirements and select freely is advanced.

(3) How to organize missions in a realistic and practicable sequence?

(a) Background discussion. The next issue involves organizing the selected missions into a satisfactory sequence. This sequence is the foundation of the exercise scenario and the heart of the overall Evaluation Plan. Although Chapter 5, ARTEP 71-2, outlines the steps involved in developing a workable, doctrinally sound flow of missions, it provides few criteria for doing so. Criteria are suggested in the Field Guide, page B-18.

(b) Field Guide remedies. We believe that mission organization is another area which can be interpreted either as a complex or an essentially common sense question. The simpler interpretation is preferable in the short run. We call explicitly on the senior planners' professional judgments about what a realistic sequence of events and missions would be. These judgments should wight the nature of available terrain, the types of units involved, and the amo amount of time estimated to be available. In addition, we believe the sequence should take account of at least the following:

- the requirement for compatible sequences if the two-battalion format is adopted.
- particular aspects of strategic/operational missions specific to the unit (e.g., likely European or Asian deployment).
- the basic capabilities and procedures for tactical simulation.⁸
- procedures and timing requirements inherent in the types of "on-line" feedback to be employed in the exercise.

Once these considerations have been incorporated in a tentative mission sequence, its "workability" should be refined by war-gaming, as recommended in 71-2. Only after a satisfactory sequence of basic missions is thus achieved should planners proceed to "factoring in" of supplementary and sub-unit missions.

The final stage in scheduling then becomes specification of the overall exercise scenario. This envelopes the mission/activity sequence by establishing a realistic initial strategic/tactical situation, and by relating the flow of exercise missions to that wider context. This general scenario must be derived from and consistent with the detailed sequence of missions/tasks prepared as outlined above.

In summary form, the main points incorporated in the Senior Command/Staff module were:

- Explicit reference to the exercise of professional judgment in mission sequencing and scenario construction.
- Provision of a checklist of the criteria discussed in the first paragraph on this page.
- Recommendation that the Chapter 5, ARTEP 71-2 procedural steps be adhered to, including war gaming.

⁸See Section C below for discussion of options and recommendations in this area. The "control plan" for the exercise is largely shaped by these factors, and is highly interdependent with specification of the mission/task sequence.

- Recommended deferral of preparation of the full scenario, including definition of the initial and subsequent situations, until the basic mission/activity sequence is adequately defined. This sequence should be used as the outline for development of the full scenario.

c. **Effective Exercise Control and Tactical Simulation Procedures.** The modern battle-field environment needs to be simulated. This section considers tactical control procedures that would achieve combat fidelity in spite of resource constraints.

(1) **Background discussion.** Actual combat is the only completely "free-play" tactical process. Short of combat, some type of control of tactical behavior is *always* necessary in training/evaluation exercises to simulate engagement "parameters" and to channel the action for evaluation purposes. Nevertheless, it must be recognized that the control measures required for tactical realism differ in kind from those required for orderly evaluation. In fact, the two types are significantly at odds in any field exercise. The trick then is to keep the amount of controlling (interference) done for evaluation purposes within acceptable bounds, while assuring the realism of the tactical process with sound combat simulation procedures.

Phase I efforts did not attempt to develop in detail a complete system of combat simulation techniques and devices. Many significant tools of this type are now available or will soon be available to field units.⁹ Instead, our short-run contribution was confined to helping field units understand how these two types of control activity differ, and how they interact in the evaluation field exercise.

(2) **Field Guide remedies.** Let's briefly consider the type of control required for evaluation first. Here units are placed in appropriate situations so that their performance on selected missions and tasks can be measured.¹⁰ This involves exercising control over the basic flow of tactical events and conditions so that each selected mission/task gets executed and evaluated according to standards provided in the T&E outlines selected.

The primary measures for providing evaluation control come from the selection and sequencing of missions in the basic Plan. During the exercise, evaluator/controller personnel simply manage the engagement to the extent required to stay consistent with the Plan, by timely injection of orders, rulings on maneuver, etc. However, control is exercised in moderation. The tactical process can be *overcontrolled* so that adequate realism is lost, thus compromising training objectives and validity of evaluation.

A prime example of this kind of *overcontrol* is a Plan with too much intricacy and rigidity in the basic scenario. This leads to too frequent evaluator/controller interventions, and thus overconstrains command and troop behavior. Over-zealousness in the application of the

⁹MILES, REALTRAIN, and a number of related engagement simulation systems and technologies are rapidly evolving to supplement conventional field exercise control capabilities, as described in FM 105-5, *Maneuver Control*.

¹⁰Chapter 2 considers the basic logical requirements for sound evaluation methodology.

Plan by eager evaluator/controllers can produce similar results. Both result in loss of spontaneity, the lack of freedom to make decisions and mistakes, insufficient opportunity for situations to develop fully in accordance with actual combat experience, etc.

The remedy has two aspects. The first is reflected in the way the basic scenario is structured and used. Here a simple principle applied: Do not build into the exercise any more evaluation control requirements or machinery than absolutely essential. Keep the Plan simple. Train evaluator/controllers to be as unobtrusive and hands-off while exercising evaluation-related control functions during the field execution.

The second aspect is to be prepared to devote a very substantial share of total assets to the creation and preservation of combat fidelity. A number of suggestions are made for simulating the combat environment of the modern battlefield. These are summarized in Figure 1-2, which abstracts basic tips for control planning from the Field Guide.

The basic requirement is to provide an Evaluator/Controller Group with adequate numbers, organization, communications, mobility, procedural guidelines and simulation devices. The same must be said with respect to the opposing force. Based on our observations of some six cases, an Evaluator/Controller Group of 25-30 officers and NCOs is sufficient for a one battalion exercise. Of this total, it is recommended that at least six (one control/simulation deputy at battalion, one with OPFOR, and four at company levels) be assigned exclusively to tactical control/simulation functions. Figure 1-2 presents schematically one approach to the organization of such an E/C Group.¹¹

Elements of guidance presented in Figure 1-1 are placed in appropriate sections of both the Senior Command/Staff and Evaluator/Controller Group Modules.

d. Recommendations for Follow-On Research. The recommended balance between control, evaluation and simulation is reflected in assignments of E/C duties, and instructions for simulation in the Field Guide. Tests will be conducted in Phase II.

¹¹ Chapter 2 deals directly and in substantially greater depth with the organization, training, and operations of the E/C Group.

Figure I-1

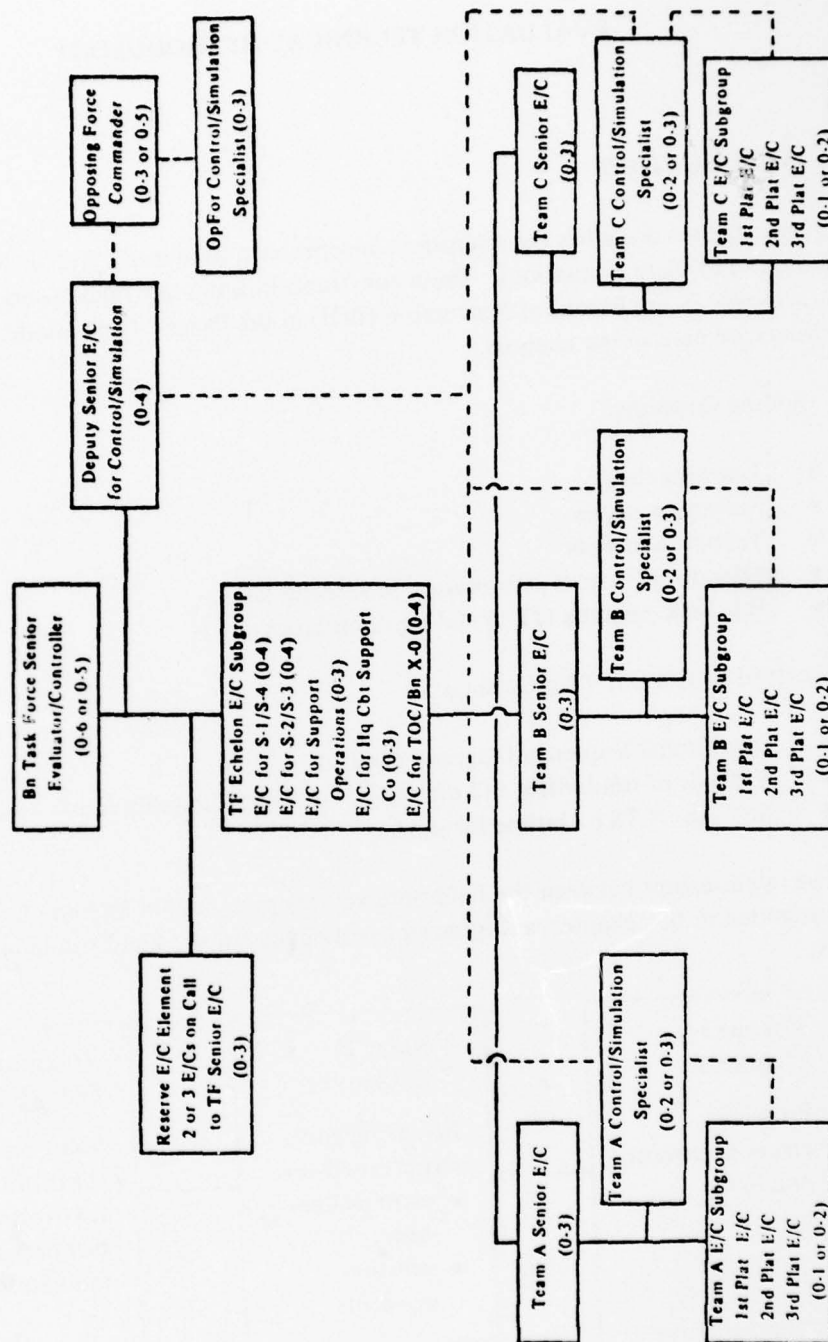
**FIELD GUIDE ABSTRACT:
TEN BASIC TIPS FOR EXERCISE CONTROL
AND TACTICAL SIMULATION**

1. Distinguish clearly between Evaluation and Control functions, both in terms of *adequately detailed procedures* and by *primary and secondary mission assignments* for all E/C Group personnel.
2. Keep control requirements to a minimum, by construction of a simple and adequately phased scenario, and by instructing evaluators in unobtrusive techniques of measurement.
3. Simulation control activities should be based on carefully formulated guidelines, which in turn reflect correct doctrinal and technical principles with respect to force ratios, weapons effects, and authentic OPFOR tactics.
4. Declaration of casualties and equipment losses should be based on codified rules expressing actual weapon capabilities, and should be applied uniformly by all appropriate E/C personnel.
5. Adequate coordination/communication capabilities must be provided to insure timely and valid rulings on representation of force ratios and declaration of casualties and equipment losses.
6. Casualty and equipment losses should be played fully for the duration of each major battalion mission.
7. Extensive use should be made of pyrotechnics and similar simulation devices, coordinated by personnel of the Control/Simulation Subgroup, and consistent with basic guidelines.
8. OPFOR should be constituted to simulate to the maximum feasible degree appropriate threat force ratios and doctrine.
9. All vehicles should be marked with distinguishing insignia, and with REALTRAIN type numbers to facilitate identification by controllers, and simulation.
10. Provision must be made for the Senior E/C, or his Tactical Control/Simulation Deputy, to monitor all important rulings, and to be prepared to resolve serious control dislocations or disputes on a timely and valid basis.

(Source: Draft Field Guide, Senior Command/Staff Module.)

FIGURE 1-2

BATTALION TASK FORCE EVALUATION/CONTROL GROUP:
PROTOTYPE ORGANIZATIONAL FORMAT



Personnel Requirements: (0-6): 0-1 (0-3): 9-13 Total All Grades: 27
 (0-5): 1-2 (0-2): 9-12 (Excluding EM)
 (0-4): 5 (0-1): 0-3

CHAPTER 2

EVALUATION TECHNICAL METHODOLOGY

2-1. INTRODUCTION

This chapter describes the scientific concepts used to identify and analyze problems in current ARTEP field evaluations. These constructs link the raw field observations to the solutions provided in the Blocks of Instruction (BOI) in the Phase I Field Guide. The following constructs are used in the analysis.

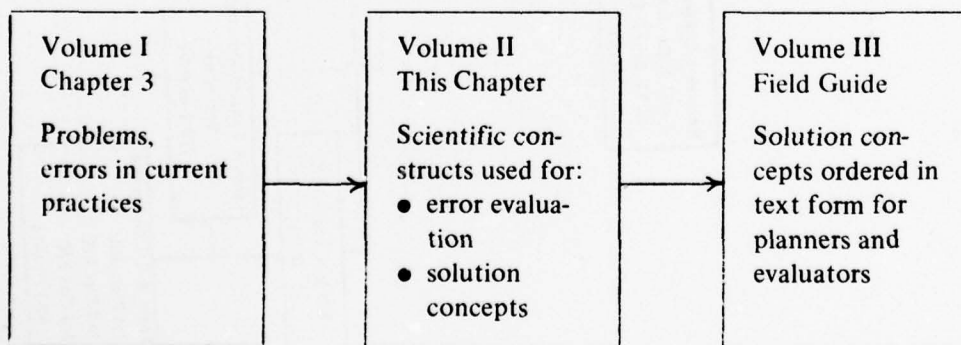
Guiding Concepts:

- Learning theory
- Systems analysis
- Tactical concepts
- "Hands-on" versus information processing tasks
- Job Task Analysis (JTA) and psychometrics

Analysis of Current T&E Outlines:

- Operational Sequence Diagrams (OSDs)
- Analysis of similarities and differences of T&E standards across echelons
- Analysis of T&E Outline format

The relationships between the field observations presented in Volume I, the analytical concepts presented in this chapter, and prescriptive guidance in the Field Guide may be shown as follows:



Each guiding concept is a method for analyzing ARTEP evaluation exercises from a different perspective, and, therefore, points out areas of deficiency in current field practice. The analyses provide guidance for improving the conduct of such exercises.

Discussion of each of these constructs is presented in the following order:

- Conceptual discussion.
- Summary of field observations and examination of T&E mission outlines.
- Field Guide applications.

The instructions and guidance included in the Field Guide represent a first attempt to integrate prescriptive guidance in a form that can be used by evaluators. The materials in the Field Guide must be validated in the second project year.

2-2. GUIDING CONCEPTS

a. Principles of Learning.

(1) Discussion. Learning of the complex tasks involved in combat unit training typically involves a sequence of steps which may be summarized as follows:

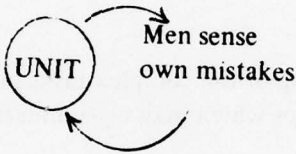
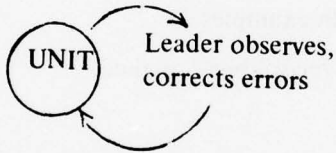
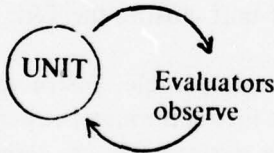
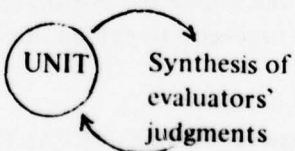
- (a) Explain.
- (b) Illustrate with examples.
- (c) Observe and record performance.
- (d) Critique.
- (e) Repeat Steps (c) and (d) or Steps (b), (c) and (d) above until the performance of the unit satisfies the T&E Standards.

Instruction and evaluation are inseparable. Instructors serve to observe, to provide feedback, and to guide repetitions. The instructor matches performance with standards, and notes deviations. Standards for effective performance are derived from doctrine. On successive repetitions, performance of the unit being trained should improve. This is the classic concept of military instruction. It emphasizes the role of the instructor in providing feedback. As we look at collective or unit training, it is necessary to extend the concept to include four feedback loops. See Figure 2-1.

In the first type, called discovery learning in REALTRAIN, the trainee recognizes his own errors. While much discovery learning naturally occurs during battalion field evaluations, it is difficult to structure conditions to provide individuals occasions for discovery learning. Nonetheless, evaluators should encourage this sort of learning by having trainees discuss what they have learned in post-mission critiques.

Figure 2-1.
LEARNING AND FEEDBACK

(Note: Evaluators assure all feedback loops work effectively.)

Type of Learning	Feedback Loops	Evaluator Responsibilities
1. Discovery learning		Evaluators encourage unit leaders to discuss events in critique; evaluators make sure <i>correct</i> tactical principles are learned.
2. From leader, peers		Evaluators encourage leaders to monitor, supervise closely. Evaluators encourage communications among peers.
3. From evaluators		Evaluators take notes, provide post-mission critique. Critique emphasizes as well self-learning, supervision.
4. From post-FEX written summary		Evaluator notes integrated, summarized after FEX by chief evaluator and reported.

Second, is feedback from supervisors and peers. Supervisors should watch and make needed corrections. But feedback should not be limited to that provided by supervisors. A well-trained unit can respond to nuances of terrain and to unexpected enemy appearances without awaiting orders from above. This need for communications among front line units on the battlefield was stressed by SLA Marshall.¹ Evaluators should encourage it.

The third source of feedback is through evaluators. It may be provided during the mission and the post-mission critique.

The fourth source is the formal evaluation summary document. It represents a synthesis of information from the E/C Group. It provides unit training managers with a basis for planning remedial training.

Self-discovery of one's own errors may well be most effective of all feedback types. That from supervisors is also important. The third and fourth forms of feedback are typical of current practice in mounting field exercises. Evaluators need to be trained so they can stimulate the operations of all feedback loops.

(2) Field observations.² Army practices observed in the conduct of battalion field exercises violate many principles of learning. Self-learning was not encouraged, and supervisors often neglected to correct errors. Typically, critiques were not held after each mission. Generally, the only form of feedback provided was the post-exercise written reports. This form of feedback is the weakest of the four types.

In one instance, field grade evaluators intervened during the conduct of the exercise to teach and to critique. A space of a post-exercise critique by one evaluator reads:

"It was apparent that they had not worked as a battalion, company, or a platoon in the field for some time. The platoon leaders were unaware of the advantages of overwatch. And this is where some four evaluators who've just recently had an opportunity to study some of this stuff decided to become instructors, and it became very valuable training experience for everybody involved. We would stop for a period of ten to five minutes, a platoon, company, and talk to them about terrain use, displacing firing units, and within a matter of fifteen minutes they were doing it, halfway properly. I think this was one of the better aspects. Over a period of four days, the battalion improved immensely, I thought . . . specifically, with regard to the maneuver of combat vehicles. It looked better toward the end than it did toward the beginning. We spend a lot of work in that area, though."

(3) Field Guide applications. Field Guide applications encourage the operation of all four feedback loops, and develop evaluator roles in their operation. The following principles help to make feedback effective.

¹SLA Marshall, *Men Against Fire*.

²*Final Report (DRAFT). Improving Army Training and Evaluation Programs (ARTEP); Methods for Unit Evaluation. Volume I. Executive Summary: Study Design and Field Research.* 31 January 1978. Pages 49, 50.

- Evaluators compare behavior they see with that which they think to be doctrinally correct. *Their judgments must be valid.* Acting as a team, they should position themselves to observe all important errors as well as effective performance. Evaluator training must assure that the feedback provided by evaluators, and by the evaluator system, is valid.
- The evaluator needs to be credible to recipient(s). This requires training evaluators as individuals, and as a team whose members need to exchange information.
- Feedback should occur as soon after the task is performed as is feasible. The more events that intervene, the less likely feedback will be effective. The Field Guide recommends a critique on completion of each mission.
- If the task is highly routinized, feedback can and should be specific. If the task is non-routine, feedback may emphasize the relevant principles with selecting actions observed in the field as examples.
- Feedback should not be given in such a way as to undermine respect of subordinates in their leaders.
- The evaluation must not evolve into a "report card" syndrome. The diagnostic character of ARTEPs must be emphasized. Guidance incorporated into the Evaluator Group module of the Field Guide addresses this issue in BOI II.

b. Systems Analytic Concepts.

(1) Discussion. The performing unit and E/C Group are two different systems. System analytic concepts apply to both. The E/C Group is responsible for the scenario, by inputting orders and intelligence directly to the battalion staff, and by controlling the OPFOR. The task of the battalion and its elements is to respond according to doctrinal rules. The evaluator system compares the performance observed against its preestablished norms for tactics and techniques. The evaluators note errors and effective performance on the T&E outline format for post-mission and post-exercise review.

In order to perform its evaluation functions, the E/C Group must perform several other roles as well. System analysis concepts can help to understand the roles of evaluators and the need for integrating them in the E/C Group system.

- The distinction between the battalion performing system, and E/C Group is not always appreciated. Each has a separate set of goals, functions and a distinct structure. A field exercise brings these two systems into continuous interaction. The interaction must be managed by the E/C Group.

- The E/C system must accomplish a number of explicitly defined functions. Group members must be trained to perform each function. Training requires establishment of SOPs, and communication procedures, so that efforts of all members of the E/C Group are integrated to accomplish the functions of the evaluator system.

(2) Field observations.³ The following field observations are pertinent to E/C systems functions.

- E/Cs received little training as individuals; almost no attention was given to training the group to perform as a team. Evaluators did not *think of* themselves as a team.
- Functions were not well-articulated. There were no apparent priorities among functions or guides on who-does-what, when two or more functions need to be performed concurrently.
- Few attempts were made to anticipate needs for communications between evaluators to evaluate activities requiring coordination between physically separate units. Preplanning for communications with OPFOR was generally inadequate.
- Members of the evaluator group were reacting to events after the fact. They were at times unprepared to perform and coordinate the complex and challenging activities logically required of a smoothly working evaluator team.
- The most critical deficiencies stem from inadequate training of E/C Groups. There was limited instruction on how E/C Group members should combine their dual evaluator/controller functions. There was a lack of training on how to make subjective judgments or on how interdependent observations and ratings should be handled. Most E/C Groups apparently failed to conduct terrain inspections, and training did not include classroom gaming of the exercise beforehand as recommended in ARTEP 71-2, Chapter V.

The battalion performs according to a mission/task logic; the E/C Group must implement an evaluation logic. Whereas the mission/task logic of the performing battalion is fairly explicitly defined in the T&E outlines, a similarly explicit specification of what we refer to as evaluation logic is lacking. It is addressed in Phase I guidance.

(3) Applications of system analysis concepts in the Field Guide. The Field Guide applies the system analytic concepts in concrete terms. Emphasis is placed in Block of Instruction (BOI) 3 and 4 on the development of all essential E/C Group functions, and their integration through planning, communication and coordination.

³*Ibid*, pages 45-49.

The BOIs develop the functions of the evaluator system in six categories:

- Control
- Simulation
- Evaluation
- Conduct of critiques
- Safety
- Administration

Although BOI 3 addresses all of these functions, the evaluation logic is defined by the first four.⁴ A detailed discussion of these four is developed in BOI 3.. These four are described below.

(a) Control. Control is the management of the exercise so as to accommodate evaluation requirements while not compromising realism. BOI 5 emphasizes the use of tactical control measures of the OPFOR, so the performing battalion can carry out its tasks. Communications among members of the E/C Group are emphasized so that actions of battalion elements can be anticipated and entanglements between forces avoided. Effective control provides for more training/evaluation time during the exercise and for more realistic simulation of combat situations.

(b) Simulation. E/C simulation creates the realistic aspects of combat during the exercise. Whereas control contributes to realism by reducing administrative interference, simulation does so through the injection of some characteristics of combat. These characteristics are the production of casualties, the firing signatures of weapons, and the effects of weapons. The Field Guide recommends the organization of the E/C Group to include Control/Tactical Simulation Deputies with each maneuver company and at the battalion level.

(c) Evaluation. Evaluation is the most critical function of the E/C Group. Evaluation requires the steps of observing, making valid judgments and recording. These lead to the performance rating of battalion elements at all levels. In BOI 4, emphasis is placed on providing instructions for valid evaluations.

The Field Guide attempts to provide E/Cs with procedures that can help them identify, and evaluate all behaviors described by rating items. Instructions provide guidance as to whether identical T&E items at battalion, company and platoon levels are made by adding platoon level ratings, or whether separate observations at successive echelons are required. Hence, E/Cs must take detailed notes as they observe. An example of a T&E outline reformatted to facilitate detailed recording is provided. Rating of unit performance is currently based primarily on the subjective judgments of evaluators acting as individuals and as a team. In order to achieve uniformity among evaluators, the Annotation Annex provides that the Senior E/C instruct the E/C Group in the interpretation of T&EO items during the E/C School. The Annotation Annex also provides rules for combining subunit or subitem ratings into overall ratings for parent units by task and mission.

⁴*Ibid.*, pages 45-49.

(d) Critiques. BOI 3 addresses the conduct of post-mission critiques. It stresses the need for critiques in the field at the conclusion of each mission. Procedural guidelines are provided for a session that will allow E/Cs to assemble their notes for critiques. The guidelines also emphasize that the critique is to be a learning vehicle. A discussion of the performer's perception of the situation and requirements of his mission is intended rather than a point-by-point comparison of his actions against a rigid school solution.

Phase II work in this area will focus on fitting these critiques into a workable schedule between missions, and on encouraging active participation by all performers.

c. **Tactical Concepts.**

(1) Discussion. The ARTEP T&E outlines are an attempt to codify the tactical activities of units over time. Given the mission, and the situations that develop, the T&E outlines should specify the best solutions, which derive from application of combat tactics. Therefore, planning for control and evaluation must take the characteristics of combat into account.

- Land combat introduces a sequence of challenge and responsive actions between antagonists. Attempting to create a realistic situation raises questions as to problem control and use of T&E outlines:
 - Deviation of problem from plan. The problem is developed based on forecasts as to how the evaluated unit will respond. However, in the progression of challenge and response, all actions cannot be anticipated in advance. When the problem departs from the plan, when should it be brought back on track? How are evaluators to bring it back on track without compromising tactical realism?
 - Implications for evaluation format. Field exercises must be planned to allow options for the evaluated unit. There is an optimal level of detail for T&E outlines. In preparing rating items, tradeoffs must be recognized between precision and detail, and providing for flexibility of response.
- Critical behaviors occur that are not covered, in the T&E outlines. Note the statements of the general officers about deficiencies in the ATTs.⁵ E/Cs should be able to accommodate action sequences other than those in the T&EOs. E/Cs should be able to detect and then to evaluate, critical behaviors that do not occur in T&EOs.⁶

⁵See page 14.

⁶As an example: A Senior Evaluator noted that in one evaluation, the Battalion had placed its field maintenance facilities for disabled vehicles within 25 yards of an intersection of blacktop roads that was the most logical point for enemy artillery registration within a 10-click area. Such an item cannot well be put in a T&E outline.

- Ground combat is characterized by operations under uncertainty. The commander cannot be sure of the enemy location, nor of the enemy response to his own actions. The lethality of modern weapons is such that the payoff often goes to the side that can resolve uncertainties most quickly. This is done by the collection and screening of intelligence. In planning exercises, determination of how best to provide intelligence, and how to evaluate battalion response is an important problem for evaluation.
- Certain broad guides need to be provided for the evaluation of tactics. For example, leaders may adopt either a bold or a conservative strategy. The success or failure of either will depend on the situation. The guide here, is that actions must be based on probabilities of enemy response. These probabilities—even though they cannot be precisely specified—become inputs to evaluation of actions. Thus, in a critique, the probable enemy actions as well as those actually played, become a proper topic for discussion.

(2) Field observations. Despite Army training doctrine and the "Report of Board of General Officers Appointed to Study Army Training Tests,"⁷ ARTEP Evaluation Plans and scenarios tended to take on "canned" or mechanical characteristics. E/C Group members must, of course, know the scenario well. This should not be true for unit leaders being evaluated. Unit leaders observed were almost always aware in advance of mission events before the evaluation was implemented. In some cases, leaders of evaluated units carried time schedules for missions and mission phases.⁸

(3) Field Guide applications for scenario and tactical flow. The following three elements of guidance are recommended:

- When designing the exercise problem, it is important to include surprises and unanticipated situations for the performing unit to react to.
- While the E/Cs need clearly structured T&EO plans, some flexibility should be allowed in the tactical flow of the field exercise. The plans should not be too rigid, nor should units be given a specific time schedule to follow.

⁷Creighton W. Abrams, Brigadier General, USA, "Report of Board of General Officers Appointed to Study Army Training Tests," 1 December 1959. Some excerpts: "The current ATT program, with its stylized drill and over-developed checklist, is believed to be an outgrowth of an academic environment and is the antithesis of the facts of life on the battlefield. Here the principles of war, which when taken together make up the Art of War, are the masters. . . . As a result, commanders who are being tested go to unusual and unrealistic measures to achieve a high score, often irrespective of tactical conditions or the exercise of sound judgment. The preparation for these tests frequently involves detailed and numerous rehearsals which are of value only in a very limited and unrealistic sense." The need for flexibility is well recognized in this report.

⁸Volume I, *Op. Cit.*, page 45.

- Instructions to the E/C Group need to emphasize communication between its members so they can anticipate unprogrammed events/actions. Members of the E/C system need guidance as to tradeoffs between flexibility and problem control.

d. **"Hands-On" and Information Processing Tasks.**

(1) Discussion. The battalion as a four-tier performing unit consists of a variety of elements whose activities must be coordinated. Performance requirements differ between elements and these differences bear on how the evaluators and evaluation team must operate.⁹

Proceeding from the crew/fire team level upward to battalion staff and commander, there are discrete changes in task types. At the team or crew level, performance consists primarily of "hands-on" tasks. These tasks are largely determined by the equipment men must operate and the terrain around them.

By contrast, the tasks of company commanders, battalion commanders and battalion staffs, differ qualitatively from those of hands-on performers. These are essentially the information processors and decision-makers. They are largely dependent on information received from the men they command. This information is highly selective, it lags the events reported, and may be of dubious accuracy. Their displays consist of maps, sketches, and often the mental picture that commanders can retain. These differences in task types have important implications for evaluation procedures and critiques.

(2) Field observations/inspections of T&EOs. T&EOs often reflect basically a Job Task Analysis approach to evaluation. This approach tends to be oriented toward hands-on tasks rather than toward problems of information handling at higher echelons. Further guidance is needed to help evaluate performance on higher level tasks.

(3) Task/echelon Field Guide applications. The difference between hands-on tasks of front line units and the tasks of supervisors of supervisors, have several implications for the tasks of the E/C Group.

- Evaluators assigned to battalion command, staff and company commanders must communicate with other E/Cs to determine whether commands given are actually reflected in coordinated activities at lower levels. Similarly, they must communicate to determine whether intelligence acquired at lower levels is communicated in a timely fashion to battalion level decision-makers.

⁹This discussion derives in part from our interview with Col. Ken Leuer, then a Brigade Commander with the 24th Mechanized Infantry Division.

- Operations of the four-tier battalion organization introduced opportunities for error which are not necessarily easy to detect. Errors may be "coupled," either up or across echelons, in that they may be joint errors by two or more decision-makers. Communications guides for E/Cs are developed to help them detect coupled errors and identify sources.
- Differences in task types have strong implications for the conduct of post-mission critiques. For "hands-on" tasks, the most logical place to conduct the critique is on the battlefield while the immediate nuances terrain and vegetation are still clear. Critiques of field actions cannot be very effective if they come down in writing from brigade or division after completion of the exercise.

This is less true of critiques for company commanders and battalion staff. They can profit from a well written critique with illustrations from maps on which the exercise was conducted.

**e. Constructs From Job-Task Analysis:
Psychometrics**

(1) Discussion. Job-Task Analysis (JTA) and psychometrics provide models that can be used to develop formats for performance evaluation. These guides are reflected in some of the terminology of T&E mission outlines.

(a) Job-Task Analysis. JTA stresses these points:

- Specify training objectives, and translate these objectives into specific activities of performing units.
- Stress detailed analysis of the activities required to accomplish goals. These activities can be described in varying levels of detail. The T&E standards represent a specification of activities to be performed.
- Plan scenario and test conditions so that they clearly require evaluated elements to exercise their combat-related options and choices. T&E outlines should be fitted to the terrain, and OPFOR played so as to bring out the key duties of performing units.
- Develop and refine relevant criteria. Criteria are of two types: process criteria, which refer to ongoing activities; and, product criteria which refer to measures of mission accomplishment. Most of the items that make up battalion T&E outlines describe activities or processes.

- Stress the objectivity of measurement. The objective of criteria, rather than subjective evaluations of performance, are readily applied to hands-on tasks, where time and motion type measures are appropriate. They are applied extensively in artillery T&E outlines. However, in tactical settings, most evaluations call for subjective judgments from evaluators. It is hardly possible to eliminate subjective judgments in evaluation of field performance. Emphasis should be placed on the identification and application of considerations relevant to the evaluation of performance.

(b) **Psychometrics.** Psychometrics place stress on formats for measurement of performance, and on development and consistent application of guidance for scoring and interpreting results. In performance measurement, psychometrics would emphasize phrasing of items so as to provide precise descriptions of expected performance, and placement of as many scale intervals along parameters that describe performance as observers can readily discriminate. Psychometrics stressed the need for establishment of explicit rules for combination of scores on individual items to provide overall performance scores.

(2) **Field observations/inspections of T&EOs.** Some implications of JTA/psychometrics concepts for evaluation of field exercises are:

- In many instances, E/C Group members did not use T&EOs to evaluate and record performance *as it was occurring*. The quality of post-mission critiques and post-exercise training diagnosis would suffer accordingly.
- Many T&EO items presume an appreciation of a variety of underlying factors which determine whether performance is satisfactory. A section of a Field Guide to provide instructions for evaluation of critical performances (for example, dissemination of orders, concentration of combat power) that apply across echelons and missions would be useful. Whether it is feasible to prepare such a section (without writing still another How-to-Fight Manual) will be examined in the Phase II study effort.
- Present doctrine and practice avoids providing rules for combining evaluator ratings. Rather, decisions as to how to combine individual observations are left to their subjective judgments. The instructions are:

"The overall proficiency rating for this mission is determined from the performance of the unit on each task, the primary training and evaluation standards, and the evaluator/trainer's subjective judgment as to whether the unit would have been successful on the modern battlefield had it performed as it did in this exercise."

(3) Field Guide applications. *If* the main purpose of the field exercise is to serve as a test, or to evaluate accountability, these current practices are unsatisfactory. If, however, the main purpose is training and training diagnosis, then failure to provide adequate rules for assimilation is less serious. By *not* translating items into numbers, the report card connotation of field tests is reduced. Individual items would provide better specific indicators of training deficiencies. Nonetheless, we feel that even for training and training diagnosis, explicit rules for combining scores are preferable to leaving this task to be performed substantially without guidance. Instructions for combining scores are provided in the Field Guide, pages B-25 to B-26, and C-54 to C-56.

2-3. USE OF CURRENT T&E OUTLINES AS EVALUATION INSTRUMENTS

The format of T&EOs can be improved so as to:

- assist planners in fitting problems to the ground.
- help instruct evaluators.
- provide a better tool for evaluators to record performance, and provide better post-mission critiques and training diagnosis.

To show why these improvements are needed and how they can be made, we have analyzed current T&EOs for three core missions—Movement to Contact, Hasty Attack and Defense, for platoons, companies and battalions. Results of our analyses appear under three headings.

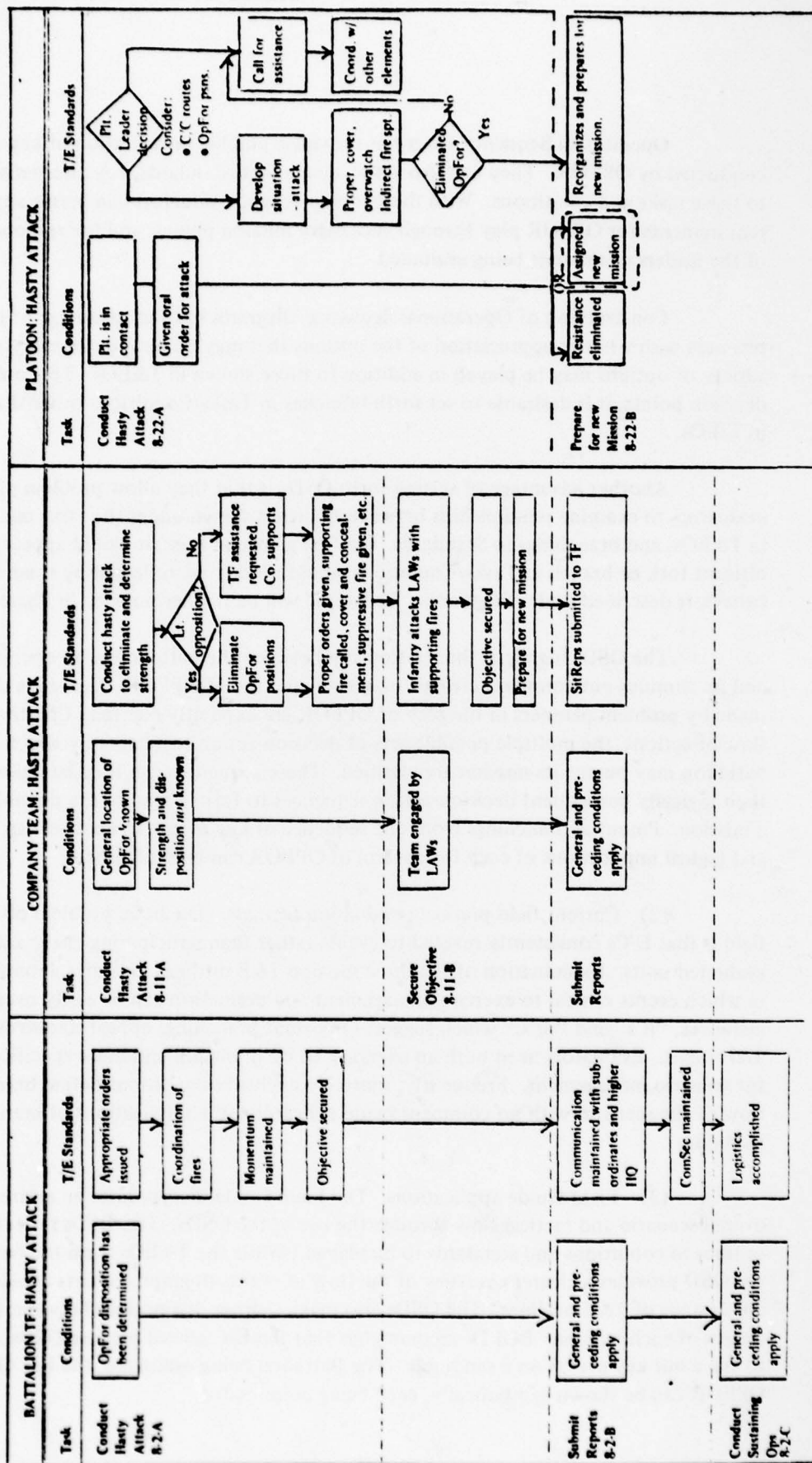
- Application of Operational Sequence Diagrams (OSDs).¹⁰
 - Analyses of inter-echelon similarities and differences among items.
 - Item/behavior parameter counts and item formats.
- a. **Operational Sequence Diagrams**

(1) Discussion. Mapping of the current T&EOs into Operational Sequence Diagrams (OSDs) is useful to both planners and the E/C Group. The planner needs to map the mission to local terrain, and to assign choices and tasks for unit leaders. This is done through operations orders and the play of OPFOR. Both planners and evaluators need an overview of the mission in skeletal form which shows critical events, and how they are ordered. The present T&EO format of certain missions tends to be cluttered. It does not give the planners or evaluators a clear picture of the essential events around which the play of OPFOR is planned, and units are evaluated. An example of OSDs for Battalion, Company and Platoons in a Hasty Attack Mission appears as Figure 2-3.

¹⁰Martin I. Kurke. "Operational Sequence Diagrams in System Design," in *Human Factors*, March 1961, Page 66.

REPRESENTATIVE OPERATIONAL SEQUENCE DIAGRAM

FIGURE 2-3



Operational Sequence Diagrams provide a graphic presentation of key actions to be conducted by OPFOR. They can also show, under T&E standards, key alternative responses to these tasks and conditions. With these, designers and evaluators can better appreciate the requirements for OPFOR play through successive mission phases, and the responses and options of the leaders of the unit being evaluated.

Construction of Operational Sequence Diagrams has the advantage of providing planners with a better appreciation of the options that may be played. In every mission, a variety of options may be played in addition to those shown in T&EOs. To emphasize key decision points, it is desirable to set forth branches in Tasks/Conditions other than those shown in T/EOs.

Another advantage of setting forth OSDs is that they allow problem planners and evaluators to examine relationships between branches shown under the tasks and conditions in T&EOs, and branchings in Standards. On first glance at least, it would appear, that a significant fork or branch in Tasks/Conditions would need to be reflected by some variations in behaviors described under Standards. This topic will be further pursued in Phase II research.

The OSDs highlight the relationship between the battalion as an operating system and its stimulus environment, created by terrain and use of OPFOR. Decisions that need to be made by problem planners in the play of OPFOR are explicitly charted. Charting the projected flow of actions, the multiple possible sets of decision-action sequences by which the performing battalion may pursue its mission are clarified. These sequences can then be linked together with their logically consequent decision-action sequences to form a map of the possible paths through a mission. Potential branchings from one sequence of key events to another can be identified and and logical implications of each for control of OPFOR can be established.

(2) Current field practice; evaluation formats. The basic problem observed in the field is that E/Cs consistently reacted to events rather than anticipating likely actions by evaluated units. Examination of the three mission T&E outlines indicates a number of instances in which events critical to exercise management and evaluation can be easily overlooked. In some instances, "if's" and "or's," which suggest important branchings appear unobtrusively within T&E items. Evaluators need both an overview of each mission and more specific instructions for scenario management. Frequently, mutually exclusive conditions define branches in the flow of the scenario with no comment made concerning the implications of branches for problem planning.

(3) Field Guide applications. The Field Guide incorporates an awareness of the overall scenario and tactical flow through the use of the OSDs. The OSDs reflect the specific options in conditions and standards to be played. While the T&EOs form the basis for the OSDs, the OSD provides a clearer overview of the flow of events in graphic charts avoiding the crowded appearance of T&E outlines. The OSDs also make sharper distinctions between the tasks or phases of each mission. BOI IV recommends that the E/C school prepare OSDs and use them to trace out key events on a sandtable. The Battalion being evaluated, the E/C Group, and OPFOR can be shown symbolically, each being color coded.

The OSDs also allow E/Cs to examine the tactical alternatives that the terrain and mission allow the performing unit. Once E/Cs have identified these alternatives, they can plan their strategies for observation keyed to the specific T&EO items, to the terrain, the tactical situation and the OPFOR play. Occasions for communication with OPFOR, and procedures for use of the E/C net(s) can also be planned.

A final application of the OSDs is in the use of a Control TOC. In battalion-size evaluation exercises, subunits are nested in the context of their parent units. The size and dispersion of battalion elements and attachments is so great that no individual E/C can keep up with events occurring at different echelons or in different subunits. Many problems arise in rating coordination among battalion elements. For example, is Company A where the battalion staff thinks it is?¹¹ Guidance has, therefore, been developed that aims at cueing of subordinate E/Cs by creating a TOC for the control system. Using the OSD as a map, the Control TOC's function would be to track the position of each E/C and the progress of each mission in relation to that envisioned by the scenario. Deviations from the master plan would then be apparent to the Control TOC and corrective action could be transmitted to subunit E/Cs.

b. Classification of Tasks and Items by Type

(1) Discussion. As documented in Issue Area III.B of Table 3-1 in Volume I (page 32), the contents of the Standards column of the current T&E outlines contain technical deficiencies. Our observations document the following specific difficulties:

- Lack of full correspondence between Tasks/Conditions/Standards items.
- Variation across missions in the detail with which behaviors are covered.
- A high degree of dependence on "professional/subjective judgment" in the interpretation of items.

Many of the T&EO items are identical at the battalion, company and platoon levels. As an example, the item "Final protective fires are fired and planned if the situation permits," occurs in the Defense T&EOs at all three echelons. This identical recurrence does not, however, necessarily mean that the required actions are also identical at all three echelons. In the above example, actions at the battalion level would involve the allocation of indirect fire support assets to line units, the decision as to whether or not to fire the final protective fires, etc. Company commanders, however, plan the precise location in which they want the final protective fires from indirect fire support batteries to fall and integrate them with the final protective fires of their organic weapons and mortars. Platoon leaders are at the greatest level of detail and plan the final protection lines of their crew-served weapons, etc.

On the other hand, some of the T&EO items that appear identically in the T&EOs of all three echelons do not represent any observable behavior at higher echelons and are clearly performed at the squad and crew level. Platoon, company and battalion performance, then, is the sum of the performances of their respective subunits. An example of this type of item is found in the React to Contact task of the Movement to Contact T&EO for all three echelons in which each echelon is to "return fire, deploy, report situation and develop situation."

What implications do these characteristics of the T&EOs have for the activities of evaluators? Do evaluators assigned to different echelons look for the same behavior, even if performance is by different people at a different echelon? Or, do identical items across echelons mean that the behavior evaluated occurs at the platoon level and that company evaluators simply sum records from their platoon evaluators? Finally, which apparently identical items actually refer to different behaviors at different echelons?

(2) Analysis of T&EO Items. The T&EO items were compared for similarities and differences across echelons (battalion, company and platoon) for three missions (Movement to Contact, Hasty Attack, and Defense). Comparisons were made across echelons within the same mission.

The initial categorization of items yielded three classes of similarity:

- 3-echelon identities: Those items appearing as standards at all three echelons. This type of item constituted 34.6 percent of the total.
- 2-echelon identities: Those items appearing as standards at only two echelons, either battalion and company, or company and platoon. These items amounted to 23.4 percent of the total.
- Non-identities: Items appearing at only one echelon. These items totaled 42.8 percent.

These identities were analyzed in terms of the following four categories of task/echelon characteristics producing the results summarized in Figure 2-4.

- Echelon-unique items: Items requiring unique information processing actions by the echelon at which they appear.
- Interactive items: Items requiring interaction between the echelon at which the item appears and another echelon. Such interactions generally take the form of dissemination of orders and the monitoring of the tactical situation through the sending and receiving of feeder reports.

- **Summative items:** Items in this category refer to those activities performed at individual, squad and crew level that are summed to provide measures of the performance of parent units.
- **Non-summative items:** This category describes activities that must be performed at various echelons, but which are not components of the counterpart activity at higher echelons. For example, the quality of lower echelon orders is not a component of the quality of higher echelon orders. Ratings of these items are not simply summed for higher echelon performance evaluation.

Figure 2-4
Cross-Tabulation of Identities by Task Type

	UNIQUE	INTERACTIVE	SUMMATIVE	NON-SUMMATIVE	TOTAL
3-Echelon	2	13	15	7	37
2-Echelon	8	10	2	5	25
Non-Identity	23	7	9	6	45
TOTAL	33	30	26	18	107
Percent	30.8	28.0	24.3	16.8	

As shown in Figure 2-4, approximately 75 percent of the 3-echelon identities fall into the interactive or summative categories. This distribution reflects the multi-echelon character of interactive and summative items. Thus, summative items at higher echelons require inputs from lower echelons before the overall parent unit performance on the item can be rated. Interactive items occur at more than one echelon and consequently, must appear in sub-unit T&EOs.

Two-echelon identities are more evenly distributed across the four types of rating items and constitute the smallest of the three identity types.

Items appearing at only one echelon, i.e. non-identities, are most heavily concentrated in the emergent property category of T&EO items. Fifty-one percent of the non-identities fall in this category. This distribution appears to support the hypothesis that at higher echelons there are unique functions that do not have counterparts at lower levels. These unique functions relate to information processing/decision-making functions that are the responsibility of command echelons.

This analysis yields two distinct results. The first has implications for the interpretation of T&EO items, while the second suggests observational procedures for use by E/Cs in the field. These considerations are embodied in the Field Guide treatment of the T&E outlines.

(3) Field Guide applications.

(a) The major problem found in the classification analysis is the presence of what appears to be emergent items in the T&EOs of more than one echelon. This anomaly appears to be

the result of imprecise phraseology. The ARTEP evaluators must, therefore, look at different aspects of a specific item depending on the level of unit being evaluated. At battalion level, evaluators judge the task organization for the mission; the selection of terrain features as objectives, battle positions, boundaries, etc.; the timely communication of intelligence and orders, the coordination of sub and supporting elements, the interpretation of reports, the general supervision of the battle. At company level, the evaluators look at a combination of supervision and of execution of the battle, such as putting out observation and security, the siting of crew-served weapons, improvement of positions. At platoon level, evaluators focus purely on execution, on the preparation of specific fighting positions, the use of concealed and covered routes, the detection of the enemy, etc. Guidance to this effect is presented in BOI

(b) The four types of rating items suggest two methods by which E/Cs may evaluate performance:

- Observations that can be made by individual E/Cs.
- Observations requiring coordinated inputs from multiple E/Cs.

1) The emergent properties items and the non-summative item types lend themselves to observation and rating by a single E/C. Emergent properties are considered to be characteristics unique to a particular echelon, generally higher echelons. They deal with the information handling process itself with less emphasis on the communication of the orders that arise from information processing. The rating items describing these behaviors can be completed on the basis of the observations of a single E/C assigned to that echelon. Non-summative items are peculiar to a single echelon at which they appear and are also observable by a single E/C.

2) Items that deal with the interactions between echelons require a plan for observation by two or more E/Cs stationed at the various levels between which the interaction occurs. Such items generally involve the handling of information from one echelon to another.

Some tasks, such as the use of cover and concealment, may not be as observable from an E/C's vantage point with the performing unit as from the opponent's side. In such cases, coordinated input from E/Cs with the OPFOR is required.

3) Some items are simply summed. These items frequently deal with hands-on tasks that occur primarily at the individual/crew/squad level. Together, they are a significant component of overall parent unit performance. Individual E/Cs at the platoon level make ratings which are then passed up to E/Cs at company and battalion levels.

c. Analysis of Formats of T&EO Items

(1) Discussion. A well written, well formatted T&E item for diagnostic use should:

- call attention to critical combat behavior,

- permit evaluators to make a permanent record of the caliber of behavior observed, as it is occurring, and
- provide the basis for a detailed critique.

Our analysis of existing T&EOs in ARTEP 71-2 showed that the above functions are not being met because too many dimensions of behavior are being compressed into single items. The separate dimensions are being confounded and there is no room for comments. For example, the item "Complete and clear orders were disseminated to all subunits in a timely manner," has four dimensions of behavior including:

- Completeness
- Clarity
- Timeliness
- Dissemination to *all* subunits

More reliable judgments could be made if these dimensions were identified for the evaluator. Moreover, these judgments could be more easily understood and interpreted by other decision makers who were not present during the exercise being evaluated.

(2) Analysis. The number of tasks, items and embedded sub-items for the three echelons by mission is shown in Figure 2-5.¹¹ This table shows two things. First, the greatest emphasis is placed on the Defense mission, which contains more items and sub-items than the other two missions combined. Second, embedding sub-items appears to be the standard practice rather than the exception. The average number of sub-items per item across echelons is five in the Movement to Contact and Attack missions, and almost three for the Defense mission. These may be taken as rough indices of item compression.

Figure 2-5
Numbers of Tasks, Standards Items and Subitems
in T&EOs by Echelon for Three Missions

(a) Movement to Contact

Echelon	Tasks	Items	Embedded Subitems	Subitems: Items	Subitems: Tasks
Battalion	5	9	51	5.67	10.20
Company	4	6	22	3.67	5.50
Platoon	3	3	26	8.67	8.67
TOTALS	12	18	99	5.50	8.25

¹¹ A rating item was considered to be any lettered entry or paragraph in the T&EO standards column. Embedded behaviors as parameters to be rated were then counted within each such item. This method of counting assumes that SAT/UNSAT ratings can be applied to each separate rating item. A strict interpretation of instructions on the cover sheet would have evaluators assign SAT/UNSAT at the task level only.

(b) Hasty Attack

Echelon	Tasks	Items	Embedded Subitems	Subitems: Items	Subitems: Tasks
Battalion	3	6	34	5.67	11.33
Company	3	5	26	5.20	8.67
Platoon	3	4	15	3.75	5.00
TOTALS	9	15	75	5.00	8.33

(c) Defense

Echelon	Tasks	Items	Embedded Subitems	Subitems: Items	Subitems: Tasks
Battalion	10	39	114	2.92	11.40
Company	9	44	141	3.20	15.67
Platoon	9	51	127	2.49	14.11
TOTALS	28	134	382	2.85	13.64

In the Movement to Contact and Hasty Attack missions, this compression makes it difficult for the evaluator to note deficiencies in individual behavior parameters: In these missions, a SAT/UNSAT rating on each item would embed an average of five individual behavioral parameters. Further, the instructions in ARTEP 71-2 on the first page of each T&EO only require a SAT/UNSAT rating for each task. Ratings at the task level would embed an average of approximately eight individual behaviors. Thus, information is lost for commanders attempting to interpret the T&E outlines.

(2) Field Guide applications.

(a) Attachment A of the Field Guide contains a proposed reformatting of the T&EOs that would help solve the confounding and information loss problem. This prototype reformatting of the battalion Movement to Contact T&EO uses the same standards now existing in the ARTEP 71-2 T&EO. Also, it provides three changes. The first of these addresses the confounding problem. Each rating parameter is extracted from its embedded prose and provided with space for a SAT/UNSAT rating. The second difference is the provision of space for general comments at the end of each task. This general comments section is provided to allow evaluators to record important behaviors not anticipated in the rating items. The final change in the reformatted version addresses the need for the evaluation of supervision by leaders. This should help evaluators encourage the operation of the second learning loop.

(b) Phase II work will include the use of reformatted T&EOs and a test of evaluator reaction to them. Further, a comparison of the value of the existing T&EO format and the proposed reformatted version in terms of useful diagnostic information is planned.

d. Annotation of T&EOs

(1) In terms of content, the T&EOs rely heavily on subjective professional judgment in the interpretation of rating items. As uniformity in such interpretations is desirable, Phase I guidance embodies three points to provide for improvement in this area:

- Identification of key terms and phrases in the three T&EOs that require interpretation.
- Specification of the responsibility of the Senior E/C to instruct the E/C Group in the correct interpretation so as to achieve consistency.
- Delineation of a procedure for uniformly integrating ratings.

The Draft Field Guide contains as part of the Annotation Annex copies of the T&EOs for the Defense missions. These are underlined to indicate those items viewed as subject to interpretative judgment. It is recommended that the E/C Group be provided with similarly annotated T&EOs during the E/C School. This procedure will help achieve a more uniform understanding of where professional judgment is to be invoked.

(2) Field Guide applications. Phase I guidance on interpretations of these items is restricted to holding the Senior Evaluator responsible for instructing the E/C Group in how he wants them interpreted. More detailed explanation of these items was not attempted in the first year of study. The feasibility of developing guidance will be examined in the second study year.

The Annotation Annex provides guidance for the uniform integration of ratings. A simple three-step integration rule that is based on the concept of mission accomplishment is provided:

- *Weigh all elements of the item equally.* For example, the item for rating the Battalion Defense Warning Order (page 8-6-2) includes evaluation criteria for "sufficient information" by "secure means" to "each company and HHC." (Three elements)
- *Determine the item rating by using the majority of SATs or UNSATs among the elements.* In the above case, if two or all three elements are performed satisfactorily, the item rating is "S," and so on. (This procedure may be used, or more stringent standards may be set.)
- *In exceptional cases, ignore the preponderant tendency if in the evaluator's judgment the unit's deficiency (or proficiency) on a single element is so extreme that it outweighs decisively the other elements in determining the unit's success on this item.* In the

above example, if the Warning Order contains "sufficient information," and is distributed by "secure means" but *is not received by several company teams*, the evaluator may decide to rate the unit's proficiency as "U" on this item as a whole.

CHAPTER 3

USE OF EVALUATION RESULTS

3-1. INTRODUCTION

Though one of the primary objectives of the ARTEP is to provide commanders with the means for identifying unit performance deficiencies, such identifications are not sufficient to insure a complete training and evaluation program. The information concerning a unit's performance must be fed back into the total collective training program (ARTEP 71-2, Chapter 4). The function of this feedback is to provide information needed by commanders to adapt the training program to the unit's needs. The effective use of the ARTEP feedback mechanisms is essential for effective training.

Given this need for effective training, there are at least two necessary conditions for the effective use of feedback:¹

- feedback procedures must be detailed and made available to the user; and
- commanders must use the feedback results in the appropriate manner.

Our observations of external evaluation field exercises revealed deficiencies in meeting these two necessary conditions. The most glaring deficiency in the field evaluations was the lack of detailing and disseminating formal feedback procedures. For example, the Letters of Instruction (LOI) from every visited field unit did not provide any means for either on-line feedback or a post-exercise E/C Group coordinating meeting to integrate ratings. With respect to use of results, again the LOIs provided no guidance as to who is responsible for how the results are used. Our analyses of such deficiencies in effective use of the ARTEP feedback mechanisms have led to recommendations in the Field Guide.

These recommendations parallel the two necessary conditions stated above. The Field Guide provides guidelines for formulating and communicating results and for incorporating results in future training activities. The next three sections deal with our analysis of these two issue areas as well as an extended discussion of two means of communicating results (on-line and written feedback).

3-2. FORMULATING AND COMMUNICATING EVALUATION RESULTS

Though our observations revealed numerous deficiencies in formulating and communicating results, five major issues seem to underlie the deficiencies:

¹Irwin I. Goldstein. Training: Program Development and Evaluation. (Monterey, California: Brooks/Cole Publishing Co., 1974).

- Defining the roles of E/C personnel in formulating and communicating results.
- Designating the appropriate target audience for feedback.
- Designating the appropriate timing for the feedback.
- Tailoring the content of the feedback to meet the needs of different users.
- Use of alternative media to enhance the feedback process.

Examination of the last issue has been deferred to Phase II when the recommended conduct of on-line critiques will be evaluated. It might be noted, however, that we did not observe use of training aids by units in the field, although several devices, such as sandtables and terrain cloth models used in After-Action Reviews, are common in Army training.

With the exception of the use of alternative media, all of the above issues appear to result from a single source: lack of a formal feedback plan. This lack of formal feedback processes in turn leads to inappropriate behavior by the unit leaders and E/C personnel. That is, given a particular situation, the unit leaders and E/C personnel are not able to consistently apply the appropriate type of feedback. Our analysis of this deficiency suggests that a formal feedback plan should be used and that as a minimum the plan should describe three types of feedback:

- On-line dialogues between performing unit leaders and evaluators.
- Informal post-exercise verbal critiques.
- Written formal evaluation reports.

The basis for this classification was established by joint consideration of our field observations and the ARTEP principles in light of three major qualities of performance feedback:

- Source of the feedback;
- timing of the feedback; and
- target audience for the feedback.²

a. **Dimensions of Feedback.**

(1) **Source of Feedback.** Our analysis revealed that the source of feedback is an extremely important determinant of effective performance feedback. Although the task, the performer, and peers and leaders are all possible sources of feedback (cf., Section 2-2a., this

²Ilgen, Daniel R., *et al.* Performance Feedback: A Review of its Psychological and Behavioral Effects. (Army Research Institute, Technical Report No. 1, February 1977.)

volume), the most dominant source of feedback in an ARTEP formal evaluation exercise are the E/C personnel. (Because of this dominance, only the E/C personnel were considered in this section of the Field Guide.) Since the usual procedure is to draw the E/C personnel from outside the unit being evaluated, this source of feedback is called an external agent. We have found that this choice of external agents as evaluators has the following major implications for the effectiveness of feedback:

- Research on performance feedback has indicated that recipients of feedback attend more closely to sources that are psychologically closest to them such as the self, task, supervisor and peers, in that order.³ Thus, an external agent is less likely to be attended to in a feedback situation.
- Feedback from these psychologically close sources is more likely to be recalled later during remedial training activities.
- With external agents, the likelihood of forgetting is also increased by the inherent time delay between performance and feedback given current procedures.
- An external agent is not as familiar with the performing unit as an internal agent which negatively influences the recipient's acceptance of feedback.

Because most of these implications detract from an external agent's effectiveness, the guidance provided in the Field Guide module for E/C Group personnel emphasises techniques for enhancing feedback credibility.

With respect to all four of the above drawbacks, the E/C personnel are urged to enhance the credibility of feedback they administer by increasing their technical competence and familiarity with the performing unit. Specifically in regards to the first drawback, techniques are recommended for focusing unit leader's attention to compensate for the lack of psychological proximity of external agents to recipients.

- It is recommended that feedback be constructive rather than negative in tone in order to avoid defensive reactions.
- E/C personnel are urged to focus unit leader's attention on critical actions described by the T&EOs prior to the simulated battle, and again during the on-line critiques.
- It is also recommended that E/Cs use a scratch pad for recording unit leader's observations. This training device has been successfully used in REALTRAIN AARs to focus participants' attention on common themes and shared perceptions.

³*Ibid.*, p. 12.

The second and third implications for the use of external agents involve decrements in recall and inherent time delays in administration of feedback. Our solution to decrements in recall during remedial training is to increase monitoring responsibilities of unit trainers. With respect to inherent time delays, it is suggested that on-line critiques be conducted between missions to decrease the time delay between performance and administration of feedback.

The fourth drawback cannot be remedied easily. However, preliminary recommendations involve the E/C becoming familiar with his assigned unit prior to the exercise, and a technique drawn from REALTRAIN AAR procedures. The technique is to conduct on-line critiques as dialogues because the E/C is not privy to all pertinent information. By allowing unit leaders to participate in the critique, the E/C is better able to assess leaders' frames of reference and avoid resistance to criticism. Our recommendation is that field units distinguish procedurally between internal and external evaluations and that the diagnostic purpose be emphasized by planning time for critiques in the Evaluation Plan/LOI.

(2) Timing of Feedback. As noted above, the timing of feedback administration is critical to effective use of the ARTEP feedback mechanisms. The current version of ARTEP 71-2 explicitly recommends that critiques be provided "immediately after or even during the exercise." However, our observations (follow-up interviews of participants of externally sponsored ARTEP exercises) frequently revealed three serious problems that decreased the diagnostic utility and learning potential of the evaluation results:

- The time delay between completion of the exercise and the feedback of information.
- The adequacy of information that was eventually conveyed to performing unit leaders.
- The lack of guidance on when and how to conduct on-line feedback.

For example, it typically took two weeks for a summary of the evaluation results to reach the battalion commander. In order for the evaluation results to be meaningfully interpreted by the unit commander, he must pair the feedback with the unit's past field performance. However, the delay in feeding back written evaluation results places a great burden on the unit commander's memory because the context or circumstances in which errors are made are readily forgotten or distorted. As a result, the unit commander's job of interpreting evaluation results becomes more difficult. To confound this issue further, literature on learning and forgetting suggests that errors in performance are forgotten more readily than correct performance.⁴ Thus, the commander may be less likely to remember those areas of performance in which remedial training is warranted. Moreover, results from studies of complex behaviors such as collective unit performance indicate that the problem may be aggravated by the possibility that those parts of the behavior sequence which are deficient may become associated with those that are correct resulting in the retainment of incorrect habits.

⁴*Ibid.*, p. 10.

With respect to the use of on-line critiques, a number of deficiencies were noted both in the planning and execution stages of the ARTEP exercise. Though, as noted above, ARTEP 71-2 recommends use of critiques during the exercise, it provides no guidance as to when to intervene. The degree of existing guidance provided in the sample of LOIs varied across units. Division A only required preparation of an evaluation "packet" which included completed T&EOs and supplementary notes. Although sub-unit evaluations and mission repetition were conducted by some evaluators until a standard was met, no systematic provision for these on-line training activities appeared in the LOI. Division B provided some guidance on the evaluation packet and post-exercise critiques in their LOI. However, there were no procedural guidelines in the LOI for the arbitrary terminations of missions or task/mission repetitions which researchers observed in the field. Division C provided more guidance in their LOI than the two foregoing units. Provision for a post-exercise verbal critique included timing and types of attendants. A significant improvement over the other two division LOIs was the inclusion of an evaluation report format which was appended to the LOI and emphasized the diagnostic utility of the results. Again, there was no provision for on-line training activities observed in the field. Guidance on conducting on-line critiques was not used because they were optional and relegated to the discretion of the Assistant Division Commander. Although Division D provided detailed exemplars of battalion and company evaluation packets, written guidance on on-line critiques was lacking. Post-exercise interviews indicated a great deal of intervention occurred during the exercise. It was concluded that this lack of definite policy and procedural guidelines for the administration of feedback led to confusion observed in the field.

This confusion during the execution stages of the exercise tended to undermine the realism and the validity of performance data collected by evaluators, and disrupted the logical flow of the exercise. For example, one division commander terminated missions arbitrarily to critique involved units over the net, which one can infer publicly humiliated unit leaders. Spokesmen of another ARTEP exercise announced they were emphasizing "big T, little e" by allowing company team evaluators to sporadically critique and instruct units and unit leaders during the tactical phase of the exercise. During the same ARTEP exercise, deliberate repetition of critique and repetition of tasks and/or missions were observed. In the majority of cases, on-line feedback was administered inconsistently and on an "ad hoc" basis. Although there was no doubt that ARTEP evaluation exercises contain some training value, the temptation to enhance or emphasize training during the exercise at the expense of evaluation must be resisted.

Given the above deficiencies in the timing of feedback, our issue resolution procedure produced guidance that emphasized the evaluation function of external evaluation exercises, while limiting on-line training activities to between-mission critiques and provision in the LOI or Evaluation Plan for referral of within-mission intervention activities to the Senior Evaluator for approval.

In order to reduce the effect of deficiencies due to time lag and to increase the quality and detail of diagnostic performance feedback, a list of feedback events is provided in the Field Guide. This list will also be discussed below under Section 3-2b. However, at this point it can be noted that time delay reduction at certain points in the ARTEP process should ensure that the appropriate information is received by the trainers of the performing unit.

(3) **Target Audiences for Feedback.** The third critical dimension of performance feedback is the target audience. Guidance in ARTEP 71-2 (Chapter 5) on appropriate target audiences is vague. The most specific guidance is provided the Senior Evaluator, who is directed to critique "selected personnel of unit." The target audience is not mentioned where general guidance for all evaluators is provided under Step 10. Based on the results of the issue resolution procedure and results from studies on REALTRAIN After-Action Reviews, it was decided to recommend that on-line critiques be conducted for unit leaders only. They in turn transmit what they learn during the critiques to their units. Not only does this obviate a potentially embarrassing situation in which leaders are criticized in front of their men, but it also reduces participants to a manageable number given the limited time available between missions and the participative nature of the critiques.

In addition to the above recommendation, it was also decided to recommend conducting separate critiques at each discrete level of command. This decision was based on two rationales. First, due to the physical dispersion of units during the play of a problem, it was reasoned that time constraints between missions prohibited assembly of all subunit leaders with their parent units. Secondly, it is a means of increasing the training value of the exercise, especially for unit leaders below company level. Phase I observations indicated relatively less activity occurred below company level because orders from higher echelons were delayed.

b. **Types of Feedback.** Our analysis suggested that three types of feedback activities should be distinguished based on the source, timing, and target audience of feedback:

- On-line dialogues between performing unit leaders and evaluators.
- Informal post-exercise verbal critiques.
- Written formal evaluation.

Using these three types of feedback activities, we categorized six feedback events which should aid in removing some of the deficiencies discussed above in section 3-2.a. The six events are distinguished from one another on the basis of differences in source and timing of feedback and the target audience for the feedback. The events are:

- Oral platoon, company and battalion on-line critiques conducted between missions.
- A battalion informal summary critique conducted following the last mission.
- A battalion evaluation team formal coordination meeting conducted the day after completion of the field exercise.
- Completed T&E outlines and supplementary notes furnished battalion and company commanders the day after the field exercise.

- Preparation of a formal written report for the performing unit commander within two weeks following the ARTEP exercise.
- Preparation of a formal written summary report to be furnished the sponsoring brigade or division headquarters.

The members of the above list of events are categorized and treated in the Field Guide as the three types of feedback activities:

- 1st event is considered an on-line dialogue and/or a post-exercise critique.
- 2nd event is a post-exercise critique.
- 4th, 5th and 6th events are considered as written formal evaluations.

The Field Guide recommends that the six feedback events be executed for all formal ARTEP evaluation exercises. We now turn to a more detailed analysis of specific recommendations for on-line and written feedback.

3-3. ON-LINE AND WRITTEN FEEDBACK

a. **On-Line Feedback.** Specific recommendations on the preparation, frequency, content and conduct of on-line critiques is provided in the E/C Group module of the Field Guide.

Recommendations on preparatory procedures for on-line critiques emphasize techniques for increasing E/C credibility and unit leaders' receptivity to feedback, and arranging for any necessary coordination among E/Cs. During the Project Guidance Workshop in Phase I, it became apparent that much work is needed in bolstering the technical competence of E/Cs. Representatives from TRADOC and FORSCOM concurred on this point. Present guidance in ARTEP 71-2 assumes E/Cs are technically competent and, therefore, never addresses the problem. We recommend that E/Cs study and thoroughly familiarize themselves with the T&E mission outlines they have been assigned and, if necessary, cross-reference appropriate FMs. We also recommend that E/Cs review their field notes prior to conducting an on-line critique and coordinate observations with the control/simulation specialist when necessary to ensure a more balanced perspective of the action. Conducting platoon and company critiques concurrently precludes the presence of company evaluators at platoon critiques. We recognize that platoon evaluators may lack significant contextual information on an event requiring communication with company evaluators, who have a broader perspective of the battle. Therefore, allowance is made in the E/C Module of the Field Guide for radio communication between evaluators for coordination purposes.

Frequency of on-line critiques has already been discussed in terms of decreasing the time lag between performance and feedback. However, two decisions were made regarding the recommendations in this section of the Field Guide. First, if the unit completes performance of one mission and proceeds to the next mission without interruption, how does this affect the contents and duration of the on-line critique? It was decided that the critique should cover performance of both missions plus an additional review of how the missions flowed into each other. Presumably, the latter would reinforce good performance. The duration of the critique may be increased in this case but not doubled due to time constraints. The second decision involved variations across echelons in when to begin and end on-line critiques. Based on Phase I field observations, we know that platoons and companies experience some down time while battalion staff plans for the next mission. Therefore, we recommend that latitude be given E/Cs as to when they begin and end on-line critiques.

Guidance on the conduct of on-line critiques is based on REALTRAIN experiences with AARs. One of the findings of HSR's previous work on AARs was that critiques tend to break down if conducted as authoritarian teaching sessions rather than as dialogues. Guidance on the content of critiques was not specific in order to allow E/Cs the degree of latitude necessary to tailor feedback to unit needs. We did recommend that all critiques cover three problem areas which are common both in REALTRAIN and ARTEP exercises.

- Control of exercise.
- Use of communications.
- Tactical execution of exercise.

During the initial phases of the issue resolution procedure, a full evaluation team summary critique was visualized at the culmination of the exercise. However, practical considerations led to an alternative option which better conforms with the expected sequence of events. The alternative option was to conduct an on-line critique of *all* missions. The choice of which option to exercise was left up to the battalion commander. The guidance provided clearly favors the second option. Several disadvantages associated with the first option are enumerated in the Field Guide.

b. **Written Evaluation Reports.** The coordination and integration of observations and ratings and the formulation of written evaluation reports will be discussed jointly. The only guidance in ARTEP 71-2 regarding coordination of observations and integration of ratings appears under Step 10 where the Senior Evaluator is directed to "consolidate the diagnostic feedback from all of his subordinate evaluators" (p. 5-9). Lacking guidance, individual units devised a variety of ways to consolidate information. In a number of instances, no coordinating meeting was held at all. In other instances, particularly Divisions A and B, the meetings emphasized sharing general impressions rather than systematically aggregating observations and ratings. In the case of Division A, heavy group pressure was evident during the mass evaluator meeting, possibly to influence individual evaluators' judgments. Division B conducted two meetings. We were informed that a full evaluation team meeting was conducted in order for company and battalion evaluators to consolidate platoon and company evaluations, respectively. Then

an informal pre-critique meeting between the Division Commander and Senior Battalion Evaluator was held to discuss interpretations of the exercise, mission-by-mission, and establish an agenda for the post-exercise critique. There was no written provision for a coordinating meeting in the LOIs of any division visited during Phase I of the project.

There is no specific guidance in ARTEP 71-2 on how to consolidate observations and ratings. Units visited apparently did not attempt to remedy the situation. Division B was the only unit to provide formal instructions in their LOI. But the instructions were a reiteration of the guidance provided in ARTEP 71-2, which is to base evaluation ratings on mission accomplishment.

To remedy this gap in existing guidance, we recommend provision be made in LOIs for the conduct of a formal E/C Group Coordination Meeting. We envision this meeting to have two purposes:

- Internal E/C Group resolution of rating problems, and
- Review of the evaluation process for future refinements.

It is up to the discretion of the battalion TF Senior E/C whether to conduct the meeting in one or two sessions, depending upon the number of problems encountered during the exercise, etc.

In our guidance, we identify two primary types of T&E items requiring discussion and resolution during the meeting. We also recommend that the E/C Group, as a collective, be responsible for identifying the probable sources of deficiencies they observed and for recommending remedial training. Finally, the sponsoring headquarters need feedback for their evaluations. It is recommended that the Senior E/C brief headquarters on the collective recommendations for refining the evaluation process.

3-4. USE OF RESULTS

The research team had but very limited opportunity to observe use of evaluation results. Therefore, we attempted to determine by phone and correspondence whether and how results were being used for remedial training. How we think results should be used and problems reasonably to be expected are summarized:

- Different echelons use results for different purposes. Potential users and their needs must be identified. For remedial training, the more specific the deficiencies are recorded, the better. Brigade and division may not need reports at this level of detail.
- Responsibilities for use of results should be delineated in advance.

- Training programs intended to remedy deficiencies need to be examined and execution monitored to assure that the remedial training does address identified deficiencies both in the plan, and in the manner in which the plan is implemented.

(In one case, battalion commanders and S-3s were charged with analyzing results, using these to plan remedial training, justifying training plans to the brigade commander, and incorporating approved revisions into the battalion Planning Calendar.)

As with any training plan, there are problems. Among problems important to scheduling remedial training subsequent to battalion field exercises are:

- Determination of the composition of the unit(s) to receive remedial training. Even given unlimited resources, it is unclear that the best mode of remedial training involves taking the entire battalion back to the field. Composition of units for retraining will depend in some part on the specific deficiencies to be remedied. Other things being equal, it would appear desirable to conduct retraining in smaller elements. Such training would be easier to schedule. Further, it should be possible to develop scenarios and establish training settings that focus on specific deficiencies identified.
- Flexibility/inflexibility of training schedules. Remedial training must take account of existing training schedules. In Phase II of the study, we need to determine to what extent existing training schedules can be modified to use evaluation results effectively.

CHAPTER 4

4-1. INTRODUCTION

This chapter summarizes the analysis of the field observations reported in Volume 1. Analytical procedures were directed toward three areas:

- a. ARTEP Systems Concepts,
- b. Evaluation Methodology, and
- c. Use of Evaluation Results.

Each of these areas was the subject of a chapter in this volume of the Final Report.

4-2. ARTEP SYSTEMS CONCEPTS

Chapter I developed the ARTEP systems concepts relevant to battalion field evaluations. These refer to broad policy issues of concern to TRADOC/FORSCOM managers and senior military planners in field units. One such issue is the need for better integration of guiding ARTEP principles and for developing guidance based on these for field units. As the field observation of ARTEP exercises showed, commanders of field units have a limited appreciation of how these principles are applied in planning and conducting training. There is also a lack of understanding of the differences between the ARTEP philosophy and concepts and those of the older ATTs.

4-3. EVALUATION TECHNICAL METHODOLOGY

Chapter II deals with the area of evaluation methodology. Constructs from several branches of science were employed to identify existing problems, and to develop possible solutions. For example, when the ARTEP exercises are considered from the perspective of learning theory, it is apparent that many basic principles are overlooked in the management of battalion evaluations. Learning theory suggests four types of feedback that can be used in field evaluations, and the things E/Cs should do to help make feedback effective.

Concepts from systems analysis also helped to define problems confronting the performing battalion as it conducts its missions, and in the operation of the E/C Group. Job Task Analysis and the comparison of battalion tasks by echelon, helps to appreciate how tasks differ between hands-on tasks and those of command information processors and decision-makers. These differences help to better define the roles of evaluators at each echelon level for comprehensive, and valid evaluations. Operational Sequence Diagrams provide overviews of each mission. They also may be used to keep the Control TOC abreast of ongoing operations. Psychometric principles suggest improvements in the formatting of T&E items. Derivations from these scientific constructs are reflected in the Field Guide.

4.4. USE OF EVALUATION RESULTS

The analysis provided in this Chapter 3 deals with the means by which corrective feedback is provided to performing units. They include on-line feedback during mission performance, and post-mission critiques. Suggestions for processing field observations and producing the formal evaluation report at the conclusion of exercises are developed.

4.5. RECOMMENDATIONS FOR FUTURE RESEARCH

Two types of follow-on work are recommended. First, are four research tasks to be accomplished in the next year. These are:

- Validate and refine the first generation Field Guide by field testing. This includes a number of subtasks. Among these are:
 - Development of instructions for evaluators as they rate performance on those items (standards) that require professional judgment.
 - Evaluation of revised formatting of standards, and their phrasing.
 - Evaluation of alternative means of combining item scores to validly reflect unit performance.
- Identify and evaluate alternative means by which battalion evaluations can be integrated into a program involving concurrent multi-echelon training.
- Determine how ratings obtained during battalion evaluations can best be utilized in remedial training.
- Update the analytic document (Volume II, this study) based on all work performed above.

Information obtained from these tasks will shed light on additional longer-range issues. These issues will require compromises managers must make between conflicting program objectives. Examples that TRADOC/FORSCOM managers may find worthy of attention are:

- a. The need to better integrate the basic concepts underlying ARTEP, and to provide guidance to field commanders on how these concepts are applied in ARTEP exercises.
- b. Whether performance-oriented evaluation is completely sufficient to evaluating the performance of decision-makers in complex information processing tasks.
- c. The issue of implementing training diagnosis versus

4-6. THE FIELD GUIDE

The analysis presented in this volume is part of a continuing developmental effort. It will be reviewed, extended and revised during the second phase of the project. Issues in the management of field evaluations described in Volume I and II are the background of input to the Field Guide. They provide direction to the substantive areas the Guide should cover. In the Field Guide, the topics are reordered and an attempt is made to provide guidance in contexts with which officers/users are familiar, and in a vernacular to which they are accustomed.